

Александр Константинов Сергей Кузнецов Николай Скворцов

Ценность ваших данных

Текст предоставлен правообладателем

«Ценность ваших данных»: Альпина ПРО; Москва, 2022

ISBN 9785206000627

Аннотация

Что такое данные и как они появляются? Как их хранить и преобразовывать? Как извлечь ценность из имеющихся информационных ресурсов и непрерывно ее повышать? Как ускорить импортозамещение? Как наладить управление данными, чтобы достойно противостоять деструктивным воздействиям? Все это и многое другое вы найдете в книге «Ценность ваших данных».

Книга состоит из двух частей. В первой прослеживается смена парадигм в отношении к данным, происходившая от первой научной революции до четвертой промышленной.

Подробно рассматриваются особенности данных как наиболее ценного актива организации и основные барьеры на пути извлечения из них ценности. Вторая часть посвящена описанию основных подходов к устранению барьеров. Анализируются ключевые области управления данными на разных этапах их жизненного цикла – от планирования до расширения возможностей применения.

Зачем читать

Данные в качестве самостоятельного суперценного актива стремительно входят в повестку дня как менеджмента и собственников компаний, так и руководителей государственных органов и учреждений. И очень важно иметь источники информации, позволяющие его осознать, научиться с ним работать и превратить в конкурентное преимущество. Предлагаемая книга – одно из тех изданий, которые позволяют получить своевременные инструменты для создания современной высокоэффективной организации и вывода своего бизнеса в лидеры рынка.

Для кого

Книга будет полезна как новичкам в вопросах управления данными, так и опытным специалистам, которые хотят углубить свои знания в этом направлении.

Сергей Кузнецов, Александр Константинов, Николай Скворцов

Ценность ваших данных

Литературный редактор В. Доценко

Руководитель проекта А. Туровская

Дизайн обложки Е. Алёнушкина

Дизайн блока Т. Саркисян

Корректоры Н. Ерохина, Н. Казакова

Компьютерная верстка Б. Руссо

© С. Кузнецов, А. Константинов, Н. Скворцов, 2022

© Оформление. ООО «Альпина ПРО», 2022

Все права защищены. Данная электронная книга предназначена исключительно для частного использования в личных (некоммерческих) целях. Электронная книга, ее части, фрагменты и элементы, включая текст, изображения и иное, не подлежат копированию и любому другому использованию без разрешения правообладателя. В частности, запрещено такое использование, в результате которого электронная книга, ее часть, фрагмент или элемент станут доступными ограниченному или неопределенному кругу лиц, в том числе посредством сети интернет, независимо от того, будет предоставляться доступ за плату или безвозмездно.

Копирование, воспроизведение и иное использование электронной книги, ее частей, фрагментов и элементов, выходящее за пределы частного использования в личных (некоммерческих) целях, без согласия правообладателя является незаконным и влечет уголовную, административную и гражданскую ответственность.

* * *

ЦЕННОСТЬ ВАШИХ ДАННЫХ



Москва
2022

Когда шла работа над книгой «Ценность ваших данных», одному из авторов случилось оказаться в экстремальной ситуации. И та выявила все действительно важные для человека качества и ценности, но в особенности продемонстрировала ЦЕННОСТЬ ВАШИХ ДРУЗЕЙ и СЕМЬИ. Именно ДРУЗЬЯМ и СЕМЬЯМ мы и хотели бы посвятить эту книгу.

Сергей Кузнецов, Александр Константинов, Николай Скворцов

Предисловие

В XXI веке много говорят о данных: эксперты спорят, стоит ли уделять правильной организации работы с ними столь много внимания или все это хайп и надумано. Авторы этой книги уже более 20 лет наблюдают за развитием различных технологий, за тем, как рождаются методологии работы с данными, в компаниях появляются специалисты по этому направлению.

Мы постарались собрать воедино исторические, организационно-методические и технологические аспекты данных, раскрыть для читателя многообразие терминов и аббревиатур. Также вас ждет рассказ о существующем инструментарии работы с данными и о трендах, которые мы наблюдаем сегодня и увидим завтра. Все это – на языке, доступном людям без специального ИТ-образования.

Мы долго шли к решению о написании этой книги, и, как бы ни откладывали это решение, наша жизнь и работа все время подвели к необходимости описания данных, которые буквально пронизывают нашу жизнь. Для нас же ситуация получилась и вовсе уникальной, потому что мы, как профессионалы в этой области, находимся, если можно так выразиться, по обе стороны проникновения данных в жизнь общества. Именно поэтому мы все-таки решились погрузиться с головой в эту работу, которая стала, неожиданно для нас, возможностью всесторонне систематизировать не только опыт и знания, но и отраслевые компетенции заказчиков и партнеров.

Еще одним важным фактором, побудившим написать книгу, стала полугодовая работа нашего коллектива над переводом и научной редактурой второго издания легендарной книги «DAMA-DMBOK: Свод знаний по управлению данными». Результаты этой работы по достоинству оценены российскими читателями, которым книга позволила получить новые и систематизировать уже имеющиеся знания, а также стать частью международного сообщества и даже целой индустрии работы с данными, говорить с ним на языке единых терминов и подходов к процессам работы с данными.

Важно отметить, что создание книги – это процесс, напрямую связанный с импортозамещением программных продуктов, когда на смену зарубежному софту приходят российские разработки. И этот процесс стал гораздо важнее в условиях санкций против России. Период с конца февраля 2022 года стал крайне тяжелым для отечественных компаний, закупивших зарубежный софт. Особенно болезненным было «расставание» с иностранными организациями, которые в течение долгих лет поставляли нам продукты в области управления данными.

Такие компании, как IBM, SAP, Oracle, объявили об уходе из России, т. е. продаж, а главное, поддержки, выполнения проектов и много другого больше не будет. Конечно, в такой ситуации все большее количество организаций начинает принимать активные меры по импортозамещению. Это движение существовало и раньше, просто сейчас, в силу понятных причин, оно ускорилося. Во многом предвидя такую ситуацию, мы постарались учесть опыт применения российского программного обеспечения в области управления данными. Уверены, наша книга будет способствовать расширению круга отечественных литературных источников по тематике управления данными и тем самым сыграет значимую роль в ускорении всех процессов, связанных с заменой иностранных программных продуктов на отечественные.

И конечно, стоит представить уважаемым читателям одного из главных героев книги. Начиная со второй ее части, мы вводим в повествование некую абстрактную корпорацию «Телеком Дубль». Это необходимо для большей связи с реальностью, с проектами внедрения, опыт в которых у коллектива авторов суммарно более семи лет. Именно с помощью «Телеком Дубль» мы будем приводить конкретные примеры, чтобы читателям было удобнее визуализировать те понятия, концепции и инструменты, о которых будет идти речь на страницах книги. Название у компании абстрактное, она нужна нам лишь для примера увеличения эффективности ее работы при внедрении различных практик в области управления данными. Соответственно, любые сходства с реальными компаниями случайны. В начале повествования компания не обладает какими-то серьезными знаниями в области управления данными. Она ведет свою деятельность на рынке телекоммуникаций, ее клиентами являются как частные лица, так и компании. Она

имеет сложную филиальную организационную структуру, информационные системы, отвечающие за различные функции, включая фронт- и бэк-офисы. Особых успехов нет. Но с каждой новой главой ситуация будет меняться! В каждой главе мы будем описывать, как рассматриваемая проблематика выглядит на примере «Телеком Дубль», что для нее значит внедрение того или иного инструмента, что получается в результате как для ИТ-ландшафта, так и с организационной и эксплуатационной точки зрения.

Ценность ваших данных... Эти три слова описывают книгу целиком. Данные, информация, знания – все эти понятия окружают нас каждый день. Но ценность данных нужно еще раскрыть с помощью грамотного управления: без этого они могут быть лишь балластом, тянущим на дно.

При обсуждении процессов управления данными часто рассматривают две цепочки – цепочку ценности данных и цепочку поставок данных. Эти обобщенные концепции настолько важны, что мы решили отразить их на обложке. Что это за цепочки, как они соотносятся друг с другом и как выстроить их таким образом, чтобы извлечь из данных максимальную ценность? Об этом мы тоже поговорим на страницах предлагаемой книги.

Надеемся, что наша работа будет интересна самому широкому кругу читателей, которые интересуются данными. А учитывая тот факт, что данные в последние годы окутывают нас все больше и больше, круг будет расширяться.

Книга состоит из двух частей.

В первой части описывается многообразие окружающих нас данных и связанных с ними технологий, рассматривается последовательность преобразований «данные – информация – знания – мудрость», обсуждаются эволюция отношения к данным в организациях и задачи, стоящие перед ними в связи с переходом к представлению о данных как ценнейшем стратегическом активе.

Во второй части анализируются особенности данных как объекта управления, описываются функции управления информационными активами в контексте построения цепочек ценности данных и цепочек их поставок, обсуждаются основные тренды будущего в области технологий управления данными.

Введение

В 2006 году британский математик и бизнесмен Клайв Хамби ввел в обращение фразу «Данные – это новая нефть»¹. Он произнес ее на саммите топ-менеджеров по маркетингу, который проводился американской Ассоциацией национальных рекламодателей (Association of National Advertisers, ANA) в Школе менеджмента им. Келлога (Kellogg School of Management).

В том же году вице-президент ANA Майкл Палмер (Michael Palmer) в статье, посвященной тезису Хамби, развил его мысль: «Данные похожи на сырье. Оно полезно, но в необработанном виде непригодно для использования. Его необходимо преобразовать в газ, пластмассу, химикаты и тому подобные вещи, чтобы создать имеющие ценность объекты, которые обеспечили бы деятельность, приносящую прибыль; аналогичным образом и данные должны быть разложены на элементы и проанализированы, чтобы они обрели ценность»².

Спустя несколько лет, в 2013 году, глава IBM Вирджиния Рометти (Virginia Rometty), выступая перед представителями американского Совета по международным отношениям (Council on Foreign Relations, CFR), несколько перефразировала эту мысль, сравнив с новой нефтью большие данные (Big Data)³.

Сравнение данных с нефтью стало использоваться еще активнее после появления в 2017 году в журнале *The Economist* статьи «Самый ценный ресурс в мире больше не нефть, а данные»⁴.

¹ Клайв Хамби (Clive Humby) – создатель успешной программы лояльности «Клубная карта» для клиентов розничной сети супермаркетов Tesco. Подавляющее большинство источников приписывают авторство этой фразы именно ему.

² https://ana.blogs.com/maestros/2006/11/data_is_the_new.html.

³ <https://siliconangle.com/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value>.

⁴ <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>.

Наконец, не менее ярко высказался авторитетный китайский ученый и эксперт Кай-Фу Ли (Kai-Fu Lee)⁵. В своей книге «Сверхдержавы искусственного интеллекта. Китай, Кремниевая долина и новый мировой порядок», вышедшей в 2019 году, он пишет: «Если данные – это новая нефть, то Китай – это новая Саудовская Аравия»⁶.

Но одновременно с ростом количества упоминаний о данных как о новой нефти возрастало и число публикаций, выдвигающих обратный тезис: «Данные – это не новая нефть»⁷. В последнее время стали появляться статьи, рассматривающие вопросы токсичности данных⁸ и даже высказывающие мнение, что данные – это «новый мусор», не имеющий особой ценности⁹.

Очевидно, ключ к разрешению противоречия – приведенные выше слова вице-президента ANA Майкла Палмера о том, что данные похожи на сырье и для того, чтобы они обрели ценность, необходима их предварительная обработка. Но следует учитывать, что это особое сырье, во многом отличающееся от нефти и других природных ресурсов. Так как же правильно организовать его обработку? Существуют ли рациональные методы работы с данными? Ответы на эти вопросы – в соответствующем отношении к данным как отдельных компаний и организаций, так и целых государств.

Последние 10–15 лет в среде специалистов, занимающихся вопросами работы с данными, последовательно формируется представление об управлении данными как активом¹⁰. Созданы профессиональные организации, развивающие это направление. В частности, функционирует Международная ассоциация управления данными (Data Management Association International, DAMA), объединившая профессионалов в области управления данными по всему миру. Она призвана собирать, систематизировать и пропагандировать прогрессивный опыт. Регулярно проводятся различные тематические форумы, появилось большое количество книг и статей на эту тему, включая вышедшее в 2017 году (и переведенное на русский язык) уже второе издание руководства DAMA к своду знаний по управлению данными (DAMA-DMBOK2)¹¹.

Что же нужно делать компании или организации, чтобы начать обращаться с данными как с активом? Базовые шаги, позволяющие обеспечивать и наращивать монетизацию данных, известны. Они изложены, например, в книге ведущего аналитика Gartner Дагласа Лейни¹² «Инфономика: информация как актив:

5 Кай-Фу Ли – экс-президент Google China, руководитель инвестиционного фонда для китайских стартапов в области искусственного интеллекта Sinovation Ventures.

6 <https://www.mann-ivanov-ferber.ru/books/sverhderzhavy-iskusstvennogo-intellekta/>.

7 В качестве примера приводим следующие публикации:

<https://hbr.org/2012/11/data-humans-and-the-new-oil>.

<https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d>.

<https://towardsdatascience.com/data-is-not-the-new-oil-721f5109851b>.

8 <https://enterpriseproject.com/article/2019/7/data-science-data-can-be-toxic>.

9 <https://expert.ru/2020/08/6/dannyie-eto-novyij-musor/>.

10 Под активом понимается имеющийся в собственности или контролируемый экономический ресурс, содержащий в себе или производящий ценность. При этом он может быть конвертирован в деньги (монетизирован).

11 DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020 (издание подготовлено при поддержке компаний «Юнидата» и BSSG).

12 Даглас Лейни (Douglas Laney) – известный специалист в области хранилищ данных и стратегического управления данными, автор модели зрелости процессов управления корпоративной информацией компании Gartner, а также один из авторов термина Big Data (большие данные).

монетизация, оценка, управление»¹³. Тем не менее в этой же книге подчеркивается, что в большинстве организаций не введена практика последовательного управления информационными активами, как это сделано в отношении материальных или финансовых активов. Сотрудники компаний продолжают действовать по старинке. В частности, сохраняется несогласованность между управлением информационными технологиями и управлением данными. В своей переписке с Лейни Джон Лэдли¹⁴ отмечал: «Пока мне приходится наблюдать, как команды разработчиков поспешно производят сотни приложений и сервисов, не принимая во внимание вопросов использования связанных с ними данных, мне будет обеспечена полная занятость».

Недооценка перехода на новые подходы к работе с данными рискованна. В 2016 году Клаус Шваб (Klaus Schwab), основатель и президент Всемирного экономического форума в Давосе, в своей одноименной книге провозгласил четвертую промышленную революцию. Революцию, которая основана на цифровизации и характеризуется дизруптивным (ломающим привычные представления)¹⁵ воздействием на утвердившиеся традиционные компании. Чтобы удержаться на плаву, компании вынуждены трансформировать себя, используя для этого цифровые и информационные технологии (рис. В1). Шваб подчеркивает, что особую тревогу в связи с осуществлением стоящих на повестке дня революционных преобразований у него вызывают два фактора, ограничивающие их позитивный потенциал. Первый – низкий уровень управления текущими изменениями. Второй – отсутствие единой последовательной концепции их осуществления. Оба фактора распространяются и на вопросы работы с данными.

Четвертая промышленная революция основана на цифровизации и характеризуется **дизруптивным (ломающим привычные представления)** воздействием на утвердившиеся традиционные компании.

Чтобы оставаться на плаву, компании вынуждены трансформироваться, используя для этого цифровые технологии, – проводить **цифровую трансформацию**.

Компания **изобретает** заново саму себя, директор по цифровой трансформации (CDTO) выступает в роли «**внутреннего предпринимателя**».

¹³ Laney D. B. (2017). Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition.

¹⁴ Джон Лэдли (John Ladley) – признанный в мире авторитет в области управления данными. Автор книг Making Enterprise Information Management (EIM) Work for Business: A Guide to Understanding Information as an Asset («Управление корпоративной информацией в интересах бизнеса: Руководство по выработке отношения к информации как к активу») и How to Design, Deploy and Sustain an Effective Data Governance Program («Как разработать, внедрить и поддерживать эффективную программу руководства данными»).

¹⁵ <https://cyberleninka.ru/article/n/gumanitarnost-v-obrazovanii-kak-otvet-na-dizruptivnye-innovatsii>.

СТРАТЕГИИ В ОБЛАСТИ ДАННЫХ И АНАЛИТИКИ — ДРАЙВЕРЫ БУДУЩЕГО

**Рис. В1.**

Основные драйверы цифровой трансформации (Gartner)

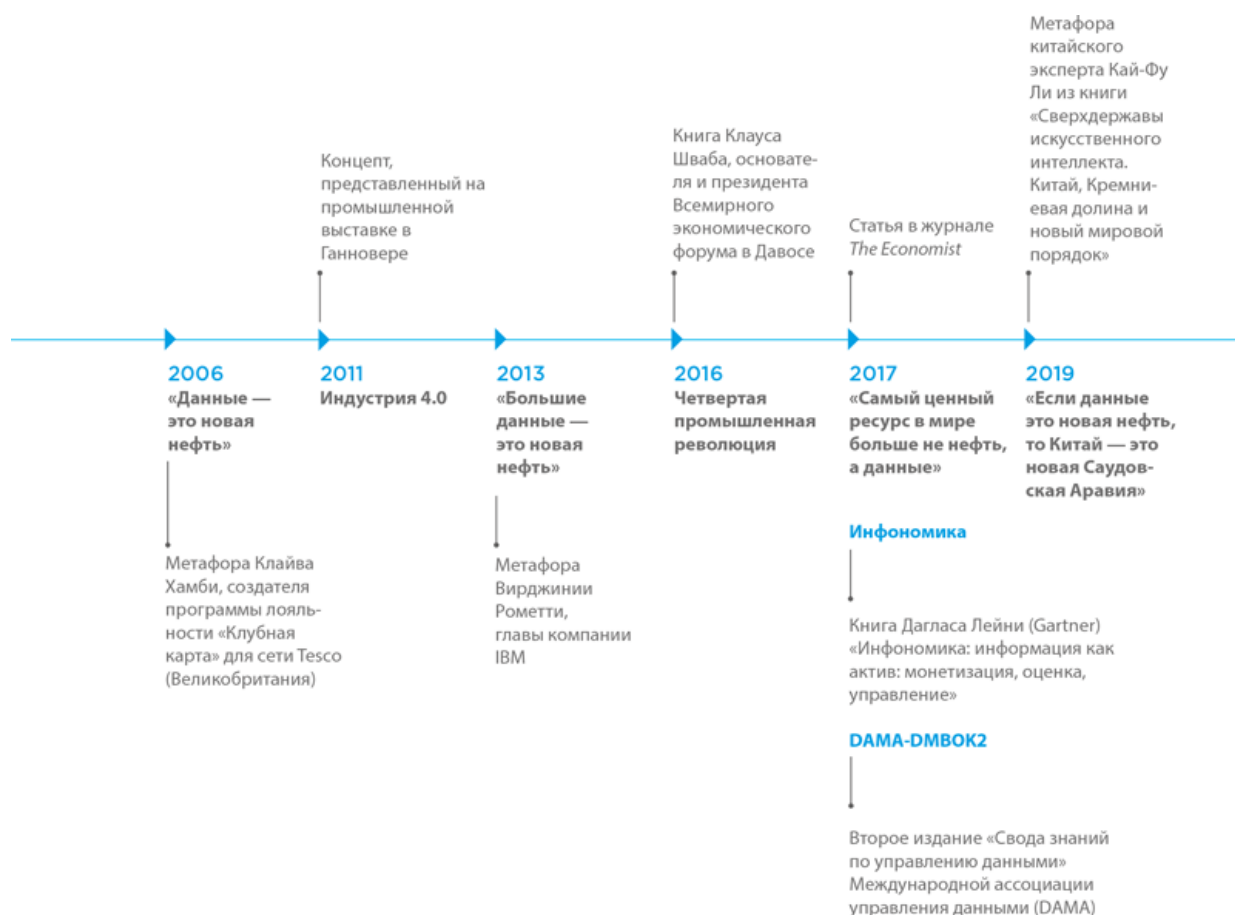


Рис. В2.

Развитие представлений о ценности данных и цифровой трансформации

Некоторые ключевые этапы развития представлений о ценности данных и цифровой трансформации начиная с начала 2000-х отражены на рисунке В2.

Появившийся во время первой президентской кампании Билла Клинтона в 1992 году популярный слоган «Это экономика, тупица» (It's the economy, stupid), подчеркивающий важность и сложность решения экономических проблем, с тех пор трансформировался применительно ко многим контекстам. В частности, в контексте цифровизации широко распространилось клише «Это данные, тупица» (It's the data, stupid)¹⁶. А чтобы привлечь еще большее внимание к необходимости комплексного учета всех аспектов работы с данными и управления ими, стала использоваться фраза «Это не просто данные, тупица» (It's not just the data, stupid)¹⁷. Важность всестороннего взгляда на данные убедительно подтверждает практика реализации проектов, связанных с уберизацией (uberization), внедрением концепции «умного города» (smart city) и другими проявлениями того, что называется цифровой экономикой (digital economy) или экономикой, управляемой на основе данных (data-driven economy).

Как наладить управление данными, чтобы достойно противостоять деструктивным воздействиям, извлечь ценность из имеющихся информационных ресурсов (превратить их из «мусора» в «нефть») и неуклонно ее повышать? Разобраться в этих вопросах руководителям и рядовым сотрудникам предприятий и организаций, не имеющим прямого отношения к сфере ИТ, порой бывает непросто. Да и у специалистов в области ИТ, как показывает наш опыт, возникает немало вопросов.

К сожалению, литературы на русском языке на эту тему еще очень мало, а информация из

¹⁶ <https://atos.net/en/blog/its-the-data-stupid>.

¹⁷ <https://www.osp.ru/news/articles/2016/25/13049787>.

многочисленных иностранных источников не всегда доступна, в том числе из-за языкового барьера. Хорошо осознавая эти проблемы, мы решили поделиться собственным накопленным опытом и видением ситуации. Надеемся также, что наша книга дополнит набор русскоязычных материалов, используемых преподавателями и обучающимися по такому недавно возникшему и стремительно развивающемуся образовательному направлению, как data literacy (грамотность в области данных).

Часть 1. Данные: Смена парадигм ***От первой научной революции до четвертой промышленной***

Глава 1. Цифровое многообразие

1.1. В чем ценность данных

Стоит только задуматься, в каком мире мы живем, и сразу обнаруживаешь, какое количество технологий нас окружает. Сейчас мы уже не представляем себе жизни без них. Мало кто задумывается, что с каждым годом динамика формирования наших потребностей во всем новом растет невероятными темпами. Даже такие крайне необходимые и уже привычные вещи, как мобильный телефон, еще 30 лет назад вызывали восхищение. Тогда пользоваться ими могли лишь избранные, сейчас же этот девайс стал постоянным спутником современного человека и уже не воспринимается как мобильное средство голосовой коммуникации. iPhone первого поколения, разработанный корпорацией Apple, поступил в продажу в 2007 году и перевернул представление о мобильном телефоне, сделав его центром коммуникаций, развлечений, устройством для получения всевозможных электронных услуг и средством платежей. Функциональные возможности современных телефонов становятся все шире. И сейчас, по прошествии 15 лет, мы себе уже не представляем, как можно жить без всего этого.

Таких примеров сотни тысяч – в самых разных областях нашей жизни. Все они стали настолько повседневными примерами нашего общения с технологиями, что мы к ним привыкли и перестали замечать.

Но несмотря на эту кажущуюся повседневность современных технологий, мировая экономика очень чутко реагирует на происходящие в нашей жизни изменения. Эта реакция проявляется повышенным интересом инвесторов в сторону технологического сегмента, который позволяет на основе человеческого интеллекта, предпринимательской смелости и современных процессов управления создавать крупнейшие мировые компании, каждая из которых может в любой момент обогнать, казалось бы, незыблемые столпы мировой экономики.

Давайте взглянем на рейтинг 100 крупнейших компаний мира (табл. 1.1)¹⁸:

¹⁸ <https://www.pwc.com/gx/en/audit-services/publications/assets/pwc-global-top-100-companies-2021.pdf>.

Табл. 1.1.

Рейтинг 100 крупнейших компаний мира (по рыночной капитализации)

Позиция	Название компании	Местоположение	Сектор экономики	31 марта 2021 г.		31 марта 2020 г.	
				Изменение по сравнению с мартом 2020 г.	Рыночная капитализация, \$ млрд	Рейтинг	Рыночная капитализация, \$ млрд
1	Apple Inc	США	Технологии	2	2051	3	1113
2	Saudi Aramco	Саудовская Аравия	Энергетика	-1	1920	1	1602
3	Microsoft Corp	США	Технологии	-1	1778	2	1200
4	Amazon.Com Inc	США	Потребительские товары вторичной необходимости	0	1558	4	971
5	Alphabet Inc	США	Технологии	0	1393	5	799
6	Tencent	Китай	Технологии	1	753	8	469
7	Tesla Inc	США	Потребительские товары вторичной необходимости	75	641	83	96
8	Alibaba Gpr	Китай	Потребительские товары вторичной необходимости	-3	615	6	522
9	Berkshire Hathaway	США	Финансы	-1	588	9	443
10	Tsmc	Тайвань	Технологии	9	534	20	235
11	Visa Inc	США	Промышленность	0	468	12	316
12	Jpmorgan Chase	США	Финансы	2	465	15	277
13	Johnson & Johnson	США	Здравоохранение	-4	433	10	346
14	Samsung Electronics	Южная Корея	Технологии	6	431	21	234
15	Kweichow Mouta	Китай	Потребительские товары	12	385	28	197
16	Wallmart Inc	США	Потребительские товары вторичной необходимости	-6	383	11	322
17	Mastercard Inc	США	Промышленность	0	354	18	243
18	Unitedhealth Grp	США	Здравоохранение	0	352	19	237
19	Lvmh Moët Hennessy	Франция	Потребительские товары вторичной необходимости	12	337	32	188

Обращает на себя внимание количество технологических компаний в первой десятке. Если быть точными, их пять, а в первой сотне – уже 30. Помимо этого, в рейтинге еще 19 компаний, занимающихся здоровьем и биотехнологиями, что также относится к технологическому сегменту деятельности. По прогнозам аналитиков, количество высокотехнологических компаний будет возрастать еще большими темпами, а одним из основных критериев успешности для компаний из других областей станет использование высоких технологий в развитии своих решений.

Экономическая важность высокотехнологических компаний уже сейчас оказывается

основополагающей для развития мировой и национальных экономик. Из 20 крупнейших компаний мира 10 (Apple, Alphabet, Amazon, Microsoft, Alibaba, Walmart, Tencent, Johnson & Johnson, Procter & Gamble, Nestlé), по мнению Boston Consulting Group, являются самыми инновационными на планете (рис. 1.1)¹⁹. Вывод прост: сегодня, чтобы стать успешной компанией, необходимо быть инновационной и высокотехнологичной независимо от отрасли деятельности. Именно это становится фундаментом для развития.

Позиция: 1–10	Позиция: 11–20	Позиция: 21–30	Позиция: 31–40	Позиция: 41–50
1 Apple	11 Siemens	21 Toyota	31 Xiaomi	41 Inditex
2 ALPHABET	12 LG	22 Salesforce	32 IKEA	41 Moderna
3 AMAZON	13 Facebook	23 Walmart	33 Fast Retailing	42 Philips
4 MICROSOFT	14 Alibaba	24 Nike	34 Adidas	43 Disney
5 Tesla	15 Oracle	25 Lenovo	35 Merck & Co.	44 Mitsubishi
6 Samsung	16 Dell	26 Tencent	36 Novartis	45 Comcast
7 IBM	17 Cisco	27 Procter & Gamble	37 Ebay	46 GE
8 HUAWEI	18 Target	28 Coca-Cola	38 PepsiCo	47 Roche
9 SONY	19 HP	29 Abbott Labs	39 Hyundai	48 Astra Zeneca
10 PFIZER	20 Johnson & Johnson	30 Bosch	40 SAP	50 Bayer

Рис. 1.1.

Самые инновационные компании 2021 г.

При всем многообразии высоких технологий в различных областях нашей жизни есть несколько ключевых технологий, на которых исследователи и разработчики строят свои открытия. Одной из групп таких технологий являются решения для работы с данными, которые не только помогают в создании новых продуктов, но и позволяют повысить эффективность действующих решений.

Одним из самых ценных активов компаний в ближайшем будущем станут данные, накопленные ими за годы деятельности: информация о клиентах, поставщиках, номенклатуре покупаемой и продаваемой продукции, о сотрудниках, оборудовании, исследовании рынков, а также статистика, публикуемая различными государственными, общественными и частными организациями. Возможность анализировать и использовать огромные массивы данных позволит государствам и организациям стать более эффективными – создать огромное количество новых направлений деятельности и совсем по-иному

¹⁹ The Most Innovative Companies 2021. The Serial Innovation Imperative. Boston Consulting Group 2021.

взглянуть на давно сложившиеся стороны жизни.

В последнее время появляется все больше и больше исследований, описывающих человека как огромный массив информации, связанной с его документами, услугами, которыми он пользуется, аккаунтами в социальных сетях, транзакциями, платежными инструментами и другими цифровыми следами повседневной жизни. Еще 100 лет назад человека, его внешность, знания, устремления, жизненный опыт, чувства, таланты, эмоции и желания изображали художники и литераторы, сейчас же цифровые следы человека заметны везде.

Как мы уже говорили, мы живем в эпоху технологий и открытий, происходящих с немыслимой скоростью, когда с момента фундаментальных открытий до повседневности технологий проходят считанные годы, а объем хранимых человечеством данных прирастает ежегодно десятками процентов. Согласно результатам исследования аналитической фирмы IDC, опубликованным в докладе «Эра данных – 2025» (Data Age 2025), в ближайшие годы основной объем данных будут производить не отдельные пользователи, а организации. На промышленность и другие сферы экономики придется до 60 % всех данных мира. Для сравнения, в 2015 году предприятия генерировали треть всех мировых данных. При этом, как утверждают авторы исследования, в будущем более важную роль будет играть качество данных, а не их количество. «Не все данные одинаково важны, а без контекста они и вовсе бесполезны. В этот период перемен лидерство будет принадлежать организациям, которые сумеют определить наиболее критичные подгруппы информации с максимальным влиянием на нужную сферу деятельности и сосредоточатся именно на них» – говорится в отчете. Пятая часть всех данных к 2025 году станет критически важной, считают исследователи. Речь идет о сведениях, от которых будет зависеть жизнь и безопасность людей, международная обстановка и мир на планете.

При этом в ближайшие годы разрыв между объемом данных, нуждающихся в защите, и реально защищенной информацией будет только расти. К 2025 году до 90 % всей информации должно быть так или иначе защищено. Авторы доклада также отмечают, что значительный объем данных будет исходить от устройств, которые окружают нас каждый день. Во-первых, к 2025 году 75 % всего населения Земли будет иметь постоянный доступ в интернет. Во-вторых, многократно возрастет количество умных гаджетов и домашних роботов, которые будут производить так называемые метаданные – служебную информацию, которой машины будут обмениваться между собой для слаженной работы. По сравнению с сегодняшним днем каждый человек будет в 20 раз чаще взаимодействовать с интернетом или с устройствами с выходом в интернет. Если сейчас среднее количество взаимодействий чуть больше 600, то к 2025 году мы будем сталкиваться с сетью 4800 раз в день²⁰.

Работа тысяч корпораций и государственных структур по созданию различного рода сервисов, предоставляющих услуги посредством интернета, уже привела к тому, что в крупных городах люди зачастую получают существенную долю государственных сервисов и приобретают различные товары и услуги за пару минут, пользуясь приложениями в телефоне. Более того, зачастую мы сталкиваемся с тем, что компании знают о наших предпочтениях и предпочтениях членов наших семей гораздо больше, чем мы сами, напоминая нам о том, какие продукты мы предпочитаем, когда необходимо запланировать очередную покупку и доставку этих продуктов к нам домой, когда необходимо сделать очередное ТО автомобиля, когда заканчиваются страховки и какие страховые продукты мы предпочли в прошлом периоде.

Все это, с одной стороны, делает жизнь легче и позволяет сосредоточиться на самом важном – семье, работе, хобби, общении с друзьями и родственниками, с другой стороны, еще больше ускоряет темп нашей жизни. Если вдуматься, мы экономим огромное количество времени на дороге, очереди, ожидании консультантов и оформлении документов в страховых компаниях, посещении всевозможных центров, предоставляющих различные государственные услуги.

Отдельно стоит отметить возможности для ИТ-специалистов, разработчиков различных приложений.

1.2. Данные для науки

Наше время породило уникальное явление, позволяющее различным научным течениям проверить и скорректировать основополагающие теории и результаты открытий прошлого. Это возможность

20 <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>.

обратиться к самому ценному и честному архиву знаний о человечестве – данным, формирующим описание человеческой природы в размерах и формах, поражающих даже самое развитое воображение. Еще никогда научному сообществу не был доступен такой объем знаний о жизни, экономике, потреблении, передвижении и любых других сферах человеческой деятельности. Вместе с тем накопленный объем информации в частных компаниях, государственных, медицинских, образовательных и социальных институтах позволяет проводить уникальные по своим масштабам и качеству результатов исследования.

Нравится нам это или нет, но информация играет все более важную роль в жизни каждого из нас, и эта роль будет становиться значительнее. Сейчас в газетах встречаются целые разделы, полностью посвященные данным. В компаниях есть группы, единственная задача которых – анализ собранных данных. Инвесторы дают десятки миллионов долларов стартапам, если те могут собрать и сохранить большие объемы данных²¹.

Знания и инструменты работы с данными позволяют формировать эффективные и точные прогностические модели даже в таких неожиданных областях, как избирательный процесс, что ранее невозможно было представить ни в социологии, ни в политологии.

Приведем несколько цитат из книги Сета Стивенса-Давидовица «Все лгут»: «...например, информация о том, кто на самом деле будет принимать участие в выборах. Больше половины граждан, которые не голосуют, говорят исследователям, проводящим опросы непосредственно перед выборами, что они намерены пойти голосовать, что искажает оценку явки, в то время как данные о поиске в Google по фразам “как голосовать” или “где голосовать” за неделю перед выборами помогут более точно предсказать, где предполагается большая активность на избирательных участках.

Можно даже узнать, за кого они пойдут голосовать. Мы со Стюартом Гэбриэлом, профессором университета штата Калифорния (Лос-Анджелес), нашли удивительную подсказку. Большой процент поисков, связанных с выборами, содержат запросы с именами обоих кандидатов. Во время выборов 2016 года, когда соперничали Дональд Трамп и Хиллари Клинтон, некоторые люди делали запрос: “выборы: Трамп – Клинтон”. Другие искали: “Клинтон – Трамп, дебаты”. По сути, 12 % поисковых запросов со словом “Трамп” включали и слово “Клинтон”. Более четверти поисковых запросов с фамилией Клинтон также содержали и фамилию Трампа. Мы обнаружили, что эти, казалось бы, нейтральные поиски могут дать нам некоторые подсказки о том, какого кандидата человек поддерживает. Как? Все зависит от порядка, в котором кандидаты появляются в запросе. Наши исследования показывают, что человек со значительно большей вероятностью поставит имя кандидата, которого он поддерживает, первым в поисковом запросе, содержащем имена обоих кандидатов»²².

«Даже если вы обманываете самого себя, Google способен узнать правду. За пару дней до выборов вы и некоторые из ваших соседей можете считать, что обязательно пойдете на избирательный участок и проголосуете. Но если ни вы, ни они не искали информацию о том, как и где проголосовать, специалисты по поиску и обработке данных вроде меня могут сказать: явка в вашем районе будет низкой»²³.

1.3. Данные для бизнеса

Отдельно следует упомянуть так называемую цифровую трансформацию – скоростное и основательное изменение жизненных процессов посредством цифровизации. Эти изменения зачастую до неузнаваемости меняют процессы на предприятиях и в государственном управлении, которым не один десяток лет, но результаты подобных трансформаций поражают. Как 30 лет назад сказал генеральный

²¹ Stephens-Davidowitz, S. Everybody lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. 2017. P. 40.

²² Stephens-Davidowitz, S. Everybody lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. 2017. Pp. 21–22.

²³ Stephens-Davidowitz, S. Everybody lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. 2017. P. 138.

директор FedEx Фред Смит (Fred Smith), «информация о посылке так же важна, как и сама посылка»²⁴.

Существует несколько ключевых факторов, играющих важнейшую роль в процессе цифровой трансформации.

- Основой решений для цифровой трансформации должны стать платформенные решения.
- Центральным местом процесса цифровой трансформации должны быть данные и процессы, позволяющие их использовать.
- Все сотрудники компании, работающие с данными, будь то юристы, менеджеры по продажам или специалисты по логистике, должны быть «владельцами данных», вовлеченными в процесс цифровой трансформации.
- Основные предметные области должны быть основой трансформации, например клиент и всевозможные данные, относящиеся к нему.
- Должен быть обеспечен цифровой профиль основных данных: чем их больше, тем лучше.
- Ключевые менеджеры должны иметь доступ к различным срезам данных для формирования нововведений и творчества.
- При достаточном уровне качества данных искусственный интеллект становится эффективным инструментарием.
- Необходимы метрики стоимости работы с данными (стоимость хранения, стоимость использования данных, эффекты от работы с данными).
- Цифровая трансформация – это процесс, увеличивающий эффективность деятельности по мере его развития.
- Цифровая трансформация без ее глубинного понимания может стать самым дорогим проектом компании и привести к ее банкротству.

Цифровые технологии можно без преувеличения назвать фактором революционного изменения всех устоявшихся бизнес-процессов и существенного повышения эффективности деятельности. Организации, которые проигнорируют необходимость цифровой трансформации или допустят ошибки на стадии проектирования этого процесса, будут неэффективными и, возможно, будут вынуждены прекратить деятельность.

Тут стоит отметить, что мультинациональные корпорации одними из первых оценили возможности использования данных как в повседневных, так и в глобальных процессах трансформаций, поняв, что они все более и более зависимы от данных организации. Вместе с тем большинство современных мультинациональных компаний не смогли бы функционировать без единой ИТ-архитектуры и единой среды обеспечения данными. Практически все процессы крупнейших компаний мира основаны на эффективных инструментах работы с данными.

Не секрет, что крупные международные корпорации в процессе своего развития зачастую создавали и приобретали компании в разных уголках мира, имеющих свою специфику в области информационных технологий и, соответственно, довольно разрозненные информационные системы относительно друг друга. В результате компания приобретала абсолютно неструктурированные и неупорядоченные данные, что существенно осложняло работу всех ИТ-систем в рамках единой организации. В этих условиях важный инструмент работы с данными в мультинациональных компаниях и организациях, повышающий эффективность деятельности компаний и учитывающий широкий портфель продуктов, – внедрение единой платформы управления данными. Вот как в общих чертах выглядит подобная единая платформа управления данными на примере крупной международной страховой группы (рис. 1.2).

Одной из самых заметных тенденций в области работы с данными в корпоративном мире в последнее время становится построение бизнес-процессов компаний на основании данных. Формируется «дата-центричный» подход, что во многом обусловлено пониманием эффекта использования данных для различных направлений деятельности компаний. Сейчас уже принимается за правило выделять и рассматривать различные информационные системы как источники данных, обогащать эти данные и формировать их ценность. Проектирование архитектуры данных становится отдельным направлением деятельности организаций. Самые важные бизнес-процессы перестраивают именно на основе этой

²⁴ FedEx Marks 40th Anniversary with Community Service, Eye to the Future <https://newsroom.fedex.com/newsroom/fedex-marks-40th-anniversary-with-community-service-eye-to-the-future>.

архитектуры (рис. 1.3).

Только после формирования архитектуры данных компании формируют функциональную и ИТ-архитектуру, описывая логическую и физическую модель работы с данными в организации. Все это приводит к тому, что фактическая функция ИТ становится вспомогательной по отношению к функции управления данными (рис. 1.4).

Чтобы оценить масштабы эффективности внедрения цифровых технологий, посмотрим на исследование, проведенное российской компанией «Юнидата». Оно показывает, что одним из самых востребованных направлений оптимизации бизнес-процессов с использованием решений для управления данными являются решения класса “Единый каталог товаров, работ и услуг” (ЕКТРУ), позволяющие повысить эффективность закупок для компаний и организаций. Те, кто пренебрегает подобными возможностями, с большой вероятностью станут неконкурентоспособными по отношению к коллегам, уделяющим должное внимание данному направлению деятельности.

Внедрение ЕКТРУ позволяет компании обеспечить прозрачность процесса закупок, вести в одном месте закупаемую номенклатуру, иметь полную информацию о товарах/услугах, их технических характеристиках, поставщиках, стоимости, аналогах, условиях и сроках гарантийных обязательств, стоимости и условиях поставки, наличии изделий на складах и в производстве и многих других параметрах. Один из следующих шагов внедрения ЕКТРУ – создание различных интерпретаций личного кабинета, позволяющего поставщикам предоставлять всю необходимую информацию о поставляемых продуктах и услугах непосредственно заказчику, а тому в свою очередь иметь актуальную информацию для обеспечения своей деятельности и оптимизации бизнес-процессов.

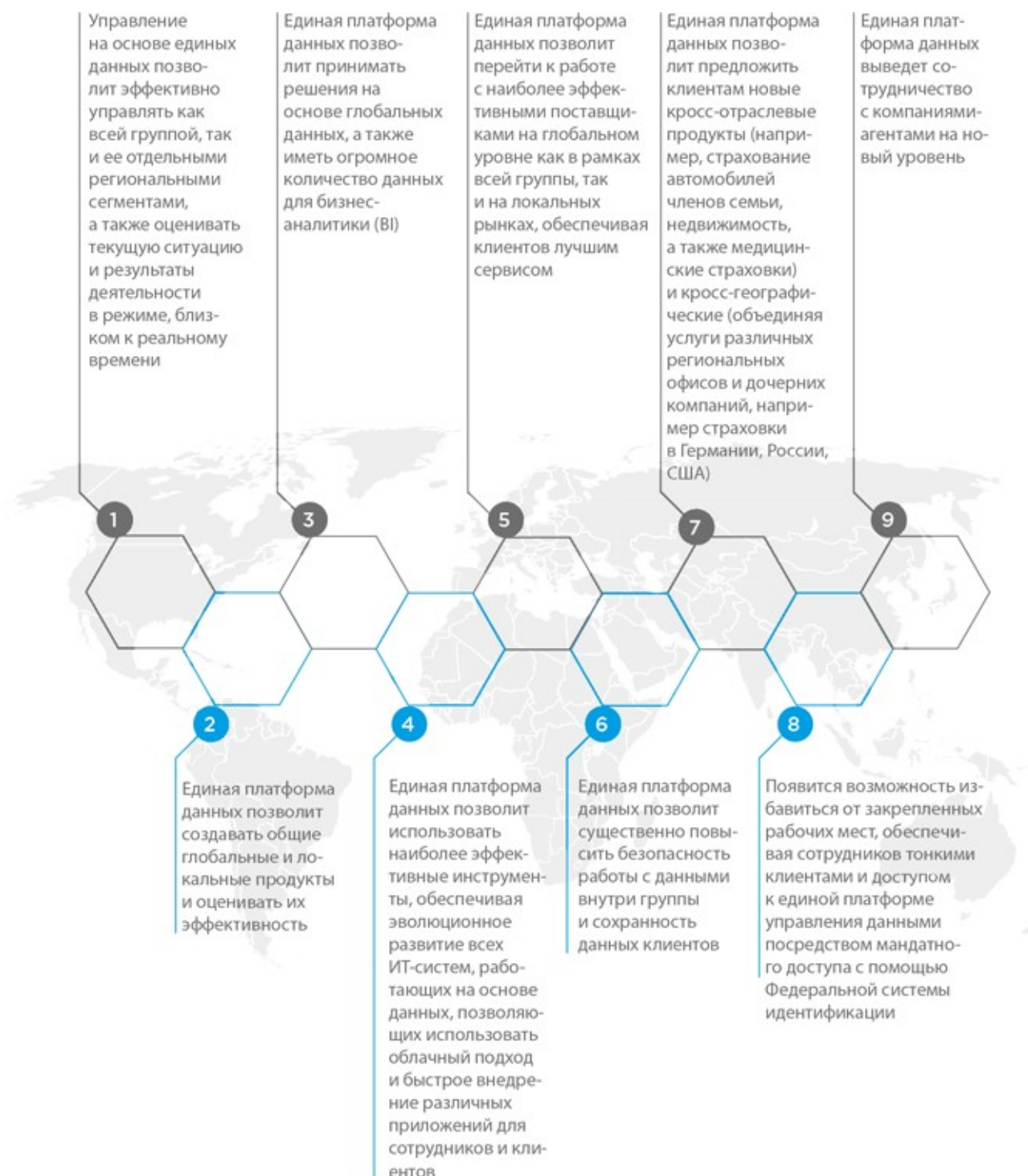


Рис. 1.2.

Единая платформа управления данными
на примере крупной международной страховой группы

**Рис. 1.3.**

Data Marketplace компании

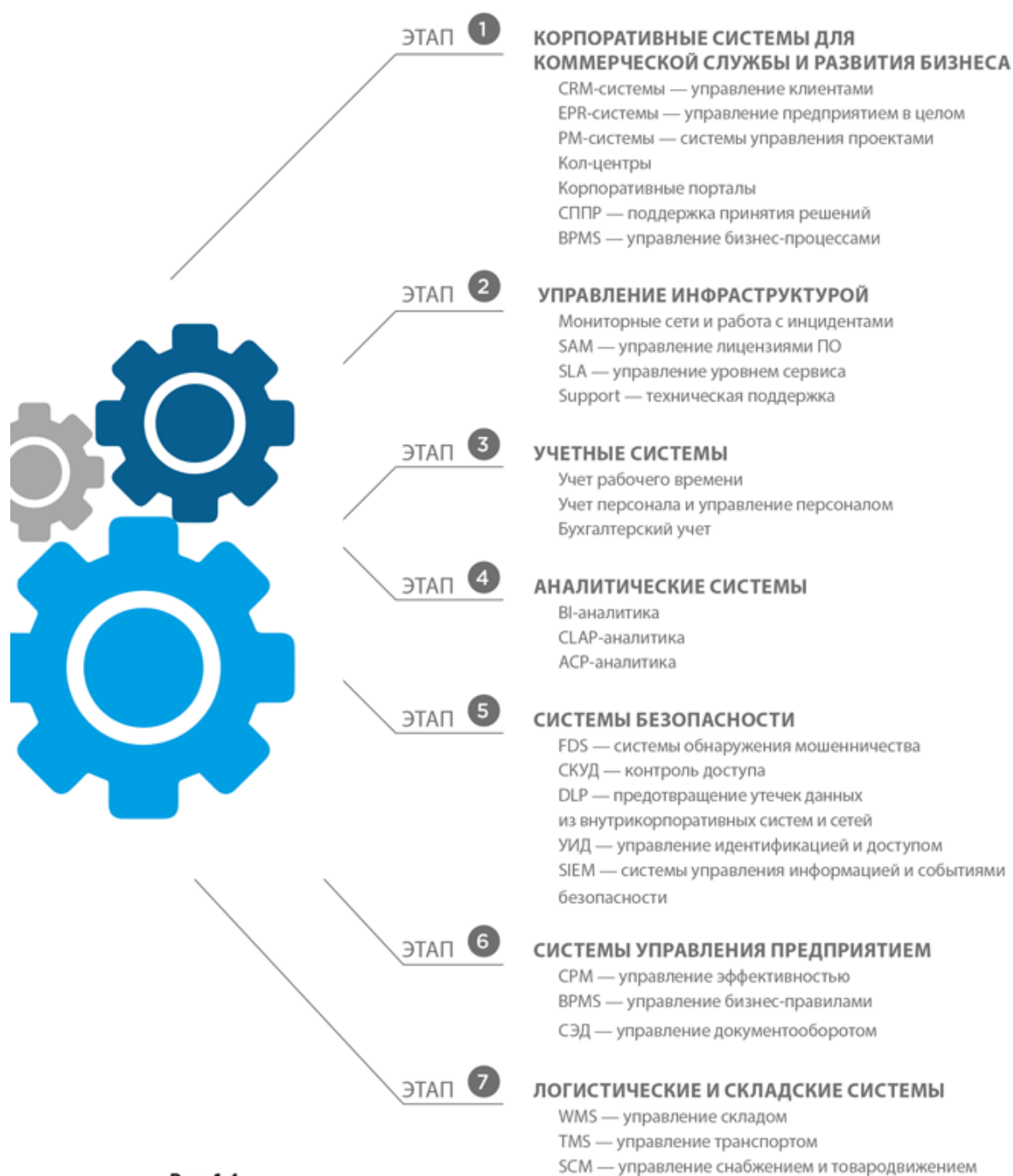


Рис. 1.4.
 ИТ-ландшафт компании

Основные результаты внедрения Единого каталога товаров, работ и услуг

- Создание и ведение параметрических классификаторов, обеспечивающих возможность формирования специализированного атрибутивного состава для каждого класса товаров, работ и услуг.
- Применение единого словаря терминов при ведении каталожных описаний.
- Возможность создания и ведения стандартных форматов описания (СФО) предметов потребления, на основании которых формируются каталожные описания конкретных номенклатурных позиций.
- Управление структурой каталога, классификаторами, опорными справочниками и другими объектами ЕКТРУ через пользовательский интерфейс платформы без привлечения разработчиков.
- Гибкая настраиваемая ролевая модель доступа к данным и функциям каталога.
- Реализация настраиваемых процессов согласования внесения изменений в каталог на основе

утвержденного регламента ведения ЕКТРУ. Обеспечение автоматизированного контроля за ходом и результатами соблюдения регламента.

- Возможность применения встроенных в платформу управления данными механизмов консолидации, дедубликации, обеспечения качества данных.

Эффекты внедрения Единого каталога товаров, работ и услуг

- Увеличение рентабельности предприятия за счет снижения складских запасов (по экспертным оценкам, от 5 до 25 %) и оптимизации резервирования материально-технических ресурсов.

- Оптимизация расходов на закупку материально-технических ресурсов за счет снижения количества избыточных заказов при закупке однотипной продукции (по экспертным оценкам, 3–5 %).

- Уменьшение ошибок в корпоративной отчетности, связанных с неактуальностью справочной информации, и, как следствие, повышение качества принимаемых управленческих решений.

- Предотвращение материальных потерь, связанных с закупкой неликвида (по экспертным оценкам, в среднем предприятия теряют из-за приобретения неликвида от 10 до 30 % от общего бюджета закупок).

- Предотвращение материальных потерь за счет экономии человеческих ресурсов предприятия благодаря грамотному и качественному обмену данными между отделами, цехами и службами (по экспертным оценкам, от 5 до 25 %).

- Достоверная информация о фактическом объеме складских запасов, статистике движения (поступления/расхода) материалов, увеличения необоснованных запасов, замораживания оборотных средств.

- Однозначная идентификация клиентов, повышение лояльности клиентов за счет наличия полной и достоверной информации о профиле клиентов, наличие информации о неблагонадежных партнерах/клиентах, повышение эффективности привлечения новых клиентов и групп клиентов за счет корректных фокус-групп.

- Повышение качества формирования закупок за счет накопленной статистики по товарам и товарным группам, поставщикам, параметрическим показателям товаров и услуг.

- Снижение затрат за счет отсутствия ошибок в адресах доставки товаров и корреспонденции.

- Сокращение поставок более дорогостоящих изделий с избыточными свойствами. Увеличение числа поставок более дешевых аналогов и заместителей, обладающих аналогичными свойствами оригинала, в том числе и при осуществлении импортозамещения.

- Снижение затрат на закупки при улучшенной консолидации закупочных партий.

- Повышение контроля в сфере закупок.

- Автоматизация всего закупочного процесса (подготовки документации для проведения закупки, определения начальной максимальной цены контракта, закупочной процедуры и последующих процессов оформления документов в строгом соответствии с закупочной процедурой).

- Снижение стоимости владения изделиями за счет подробного описания состава изделия и применимости ремонта и использования комплектующих при ремонте с учетом сроков гарантии каждой детали.

1.4. Данные для торговых сетей

Дальнейшее использование данных и технологий работы с ними изменит не только бизнес-процессы, но и сами отрасли.

Работа с данными позволит существенно повысить эффективность бизнес-процессов, снизить стоимость продукции при использовании унифицированных данных по продукции, создать единообразие продукции, увидеть наличие данных по аналогам и параметрической информации о товарах, статистику о закупаемых изделиях и наличие продукции на складах, объединять данные из информационных систем, оперирующих информацией о гарантийных обязательствах, сроках ремонта и обслуживания изделий.

Вместе с тем в будущем большинство организаций будут вынуждены кардинально перестроить свою деятельность.

Уже сегодня большой процент заказчиков покупает товары повседневной необходимости через интернет-магазины или приложения, что приводит торговые компании к необходимости оптимизации

(зачастую закрытия) торговых сетей или отдельных торговых точек. Это также связано с тем, что покупатели довольно консервативны в своих предпочтениях и приобретают одни и те же товары, потребительские свойства которых им известны. Можно с уверенностью сказать, что спектр товаров, приобретаемых онлайн, будет расширяться. Производителям и торговым сетям эти изменения позволят иметь более точную информацию о своих клиентах, их предпочтениях и пожеланиях.

Тенденция распространится на товары более длительного использования, например одежду, электронику, автомобили. Это приведет к изменениям работы торговых сетей и их взаимоотношениям с производителями товаров, появлению новых экономических схем в сфере торговли. Модель работы изменится: в торговых сетях появятся нескольких флагманских магазинов, основной целью которых будет не офлайн-продажа, а знакомство покупателей с товарами и их потребительскими свойствами. Указанная модель обеспечит огромное преимущество торговым сетям, которые первыми придут к такому формату и фактически заставят остальные сети последовать своему примеру, поскольку экономия на капитальных и операционных затратах по сравнению с классическими торговыми сетями будет гигантской благодаря отсутствию затрат на аренду помещений, оборудования, коммунальные платежи, логистику, хранение товаров, зарплаты персонала в большом количестве торговых точек и других издержек. Эта модель уже получила широкое распространение: покупатели посещают магазины/салоны исключительно чтобы увидеть искомое воочию и получить консультации о потребительских свойствах и функционале товаров офлайн, но приобретают понравившийся товар в интернет-магазине, где им предлагают более выгодную стоимость и удобство доставки.

«В середине января 2020 года производитель аудиотехники Bose объявил о закрытии 119 магазинов по всему миру. Причиной такого решения Bose назвал тот факт, что наушники, колонки и другие продукты компании все чаще приобретаются через интернет. Учитывая резкий переход к онлайн-покупкам на определенных рынках, Bose планирует закрыть все розничные магазины в Северной Америке, Европе, Японии и Австралии в течение нескольких месяцев. Однако в других частях мира магазины Bose останутся открытыми, включая около 130 магазинов, расположенных в Китае и Объединенных Арабских Эмиратах, Индии, Юго-Восточной Азии и Южной Корее»²⁵.

По этому пути уже довольно успешно идут компании и организации, предоставляющие различные услуги, которые заказчики приобретают исключительно онлайн.

Успешность деятельности этих организаций будет обеспечиваться эффективностью используемых ими ИТ-решений и в первую очередь решений по управлению данными, а традиционные для торговых сетей и предприятий затраты на аренду помещений, оборудования, стоимость персонала, оборачиваемость продукции в каждой торговой точке, перемещение товаров по цепочке «производитель – склад – магазин – покупатель» станут незначительными.

По нашему мнению, подобный формат работы торговых сетей сделает более эффективным и прозрачным механизм взаимодействия с государством, обеспечит на более качественном уровне администрирование уплаты различных налогов, акцизов, таможенных пошлин и других платежей в бюджеты государства различных уровней.

Необходимо также учитывать тот эффект, который на процессы цифровизации оказали события, связанные с COVID-19. По мнению многих специалистов, это существенно ускорило трансформацию различных индустрий. Ритейл стал одной из самых динамичных отраслей в переходе на новые цифровые технологии и процессы.

За несколько месяцев пандемии торговые сети сгенерировали огромный объем данных о своих клиентах (которые практически мгновенно из офлайна перешли в онлайн), что позволило сформировать полноценный профиль клиента. Благодаря всем этим изменениям покупатели могут анализировать свои предпочтения, использовать службы доставки, экономя время, а также более четко планировать покупки. Ритейл-компании, перестроившие свои бизнес-процессы на основе технологий работы с данными, обеспечили себя полной информацией о своих клиентах, их местах пребывания, средствах и технологиях оплаты товаров, предпочтительных программах лояльности, объемах потребления, составе семьи, предпочтениях в различных товарных группах и других данных, позволяющих существенно повысить эффективность взаимодействия с покупателями.

²⁵ <https://www.businessinsider.com/bose-stores-closing-us-brick-mortar-2020-1>.

Торговые сети уже довольно давно используют технологии работы с данными: «Например, Walmart использует данные о продажах во всех своих магазинах, чтобы знать, какие продукты следует пока отложить. До урагана Фрэнсис – разрушительного шторма, обрушившегося на юго-восток США в 2004 году, – компания Walmart подозревала (и совершенно справедливо), что, когда город переживет удар стихии, покупательские привычки людей могут измениться. Эксперты компании изучили данные по продажам после предыдущих ураганов, стараясь понять, что именно люди, возможно, захотят купить. Какой товар оказался самым популярным? Клубничное печенье. За несколько дней до урагана этот продукт продается в семь раз быстрее, чем обычно»²⁶.

1.5. Данные для государства

Сейчас даже сложно предположить, насколько управление данными может быть эффективным для государства, объем сервисов которого возрастает с каждым днем, они становятся гораздо удобнее, позволяя гражданам не только получать услуги с невиданным ранее качеством и скоростью, но и открывать иные возможности для общения государства с гражданами, а также давая возможность оперировать качественными данными для принятия управленческих решений.

С точки зрения архитектуры подобных решений необходимо рассматривать совершенно новый подход, позволяющий в дальнейшем производителям, разработчикам и различным новым функциональным заказчикам обеспечить единую платформу для построения и расширения текущих конфигураций (рис. 1.5).

Возможность обеспечения цифровых сервисов для избирателей позволит предоставить всем сторонам, участвующим в процедуре выборов, новое качество процесса, его мониторинга и прозрачности. Цифровые технологии позволят обеспечить эффективный и качественный механизм формирования реестра избирателей (в электронном виде) со всевозможными правилами качества (создание «золотой записи» избирателя для однозначной идентификации гражданина [рис. 1.6]), поддержку ведения цифрового регистра избирателя с возможностью отслеживания «жизненного цикла» избирателя (рис. 1.7), сопоставление необходимых подборок по тем или иным параметрам выбора, возможность голосования на федеральных, региональных и муниципальных выборах и референдумах по месту нахождения за пределами региона в онлайн-формате (через электронное заявление об изменении участка), возможность избирательным комиссиям и наблюдателям осуществлять онлайн-мониторинг за подготовкой, проведением и подведением итогов выборов, а также предоставление им корректных и качественных аналитических отчетов.

²⁶ Stephens-Davidowitz, S. Everybody lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. 2017. P. 95.



Рис. 1.5.

Общая архитектура государственной системы управления данными

С помощью инструментария работы с данными членами избирательных комиссий обеспечиваются прозрачность работы со всем массивом информации, корректность и прозрачность избирательного процесса (рис. 1.8).

Цифровые избирательные сервисы также обеспечат эффективную коммуникацию с каждым избирателем и позволят предоставить ему правдивую, полезную и персонифицированную информацию о выборах, а также всевозможную справочную информацию (рис. 1.9).

Вместе с тем качественная и полная информация об участниках избирательного процесса позволит предоставить цифровые сервисы в новом удобном и персонифицированном формате. Одним из цифровых сервисов избирательного процесса, существенно повышающих удобство участия в выборах, станут различные интерпретации личного кабинета (рис. 1.10–1.14).



Рис. 1.6.

Создание «золотой записи» избирателя



Рис. 1.7.

«Жизненный цикл» избирателя

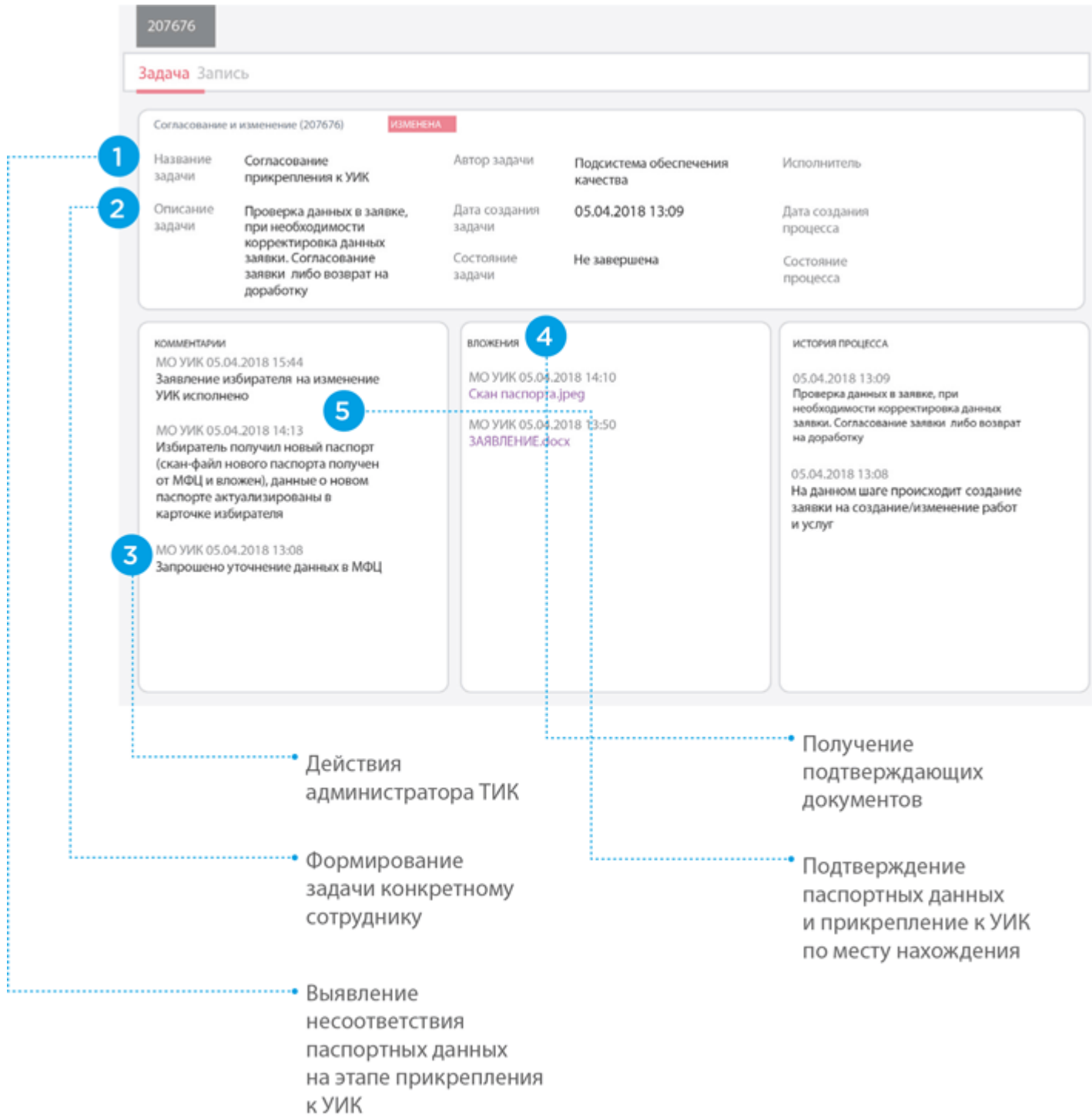


Рис. 1.8.
Анализ расхождений учетных данных избирателя

В этом разделе также хотелось бы поделиться с читателями еще одним примером крайне эффективного механизма взаимодействия государства и общества. Это, по нашему мнению, именно тот случай, когда обе стороны получают огромные выгоды от сотрудничества. Речь идет о создании электронной трудовой книжки. Чтобы избежать сомнений в важности инструментария управления данными, приведем некоторые цифры по данному направлению, которыми оперирует Пенсионный фонд РФ.



Персонализированная информация для авторизованного избирателя

- Информирование —
СМС, электронная почта, уведомления
- Таргетированная информация
- Интерактивная коммуникация

Рис. 1.9.

Коммуникация с избирателем

- 44,2 млн пенсионеров
- 154 млн актуальных лицевых счетов в системе персонифицированного учета
- 22 млн получателей социальных выплат
- 8,5 млн получателей материнского (семейного) капитала
- 8,3 трлн рублей выплачено в 2018 году в виде пенсий и других социальных выплат

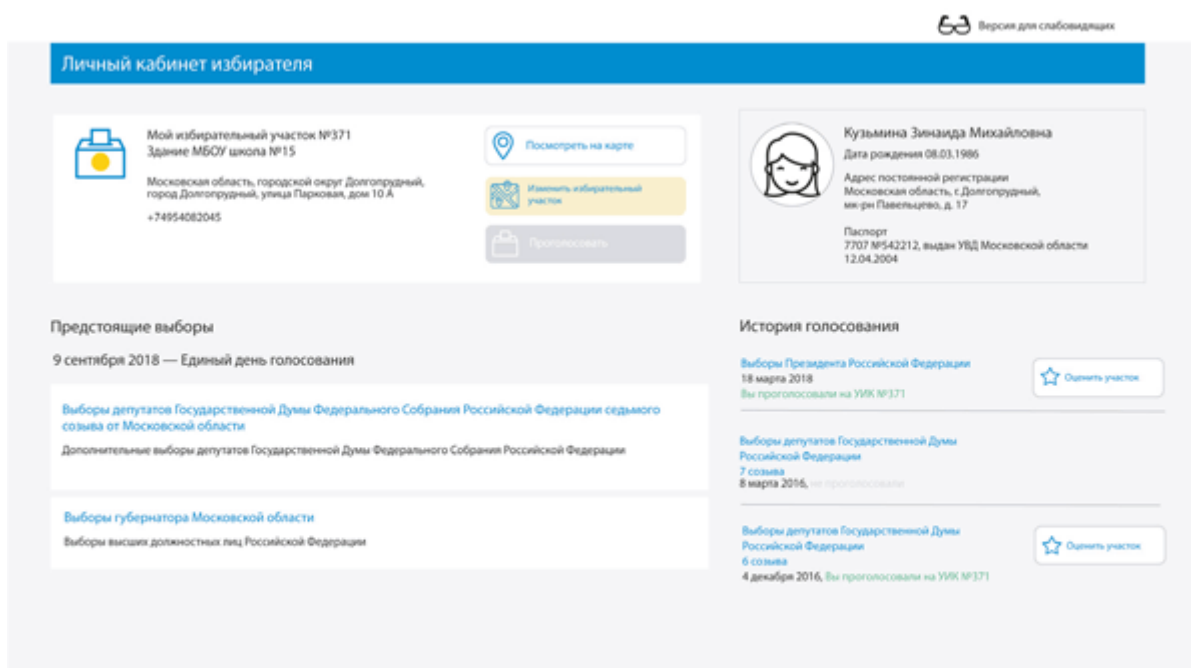


Рис. 1.10.

Личный кабинет избирателя

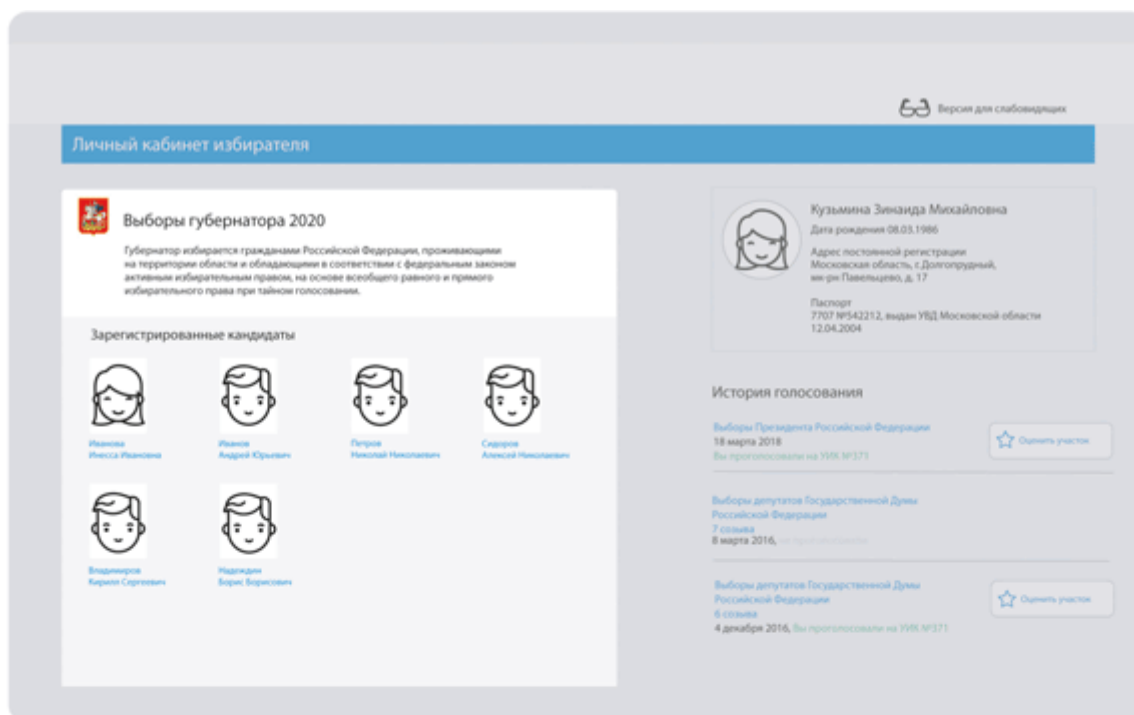


Рис. 1.11.

Личный кабинет избирателя (информация о выборах)

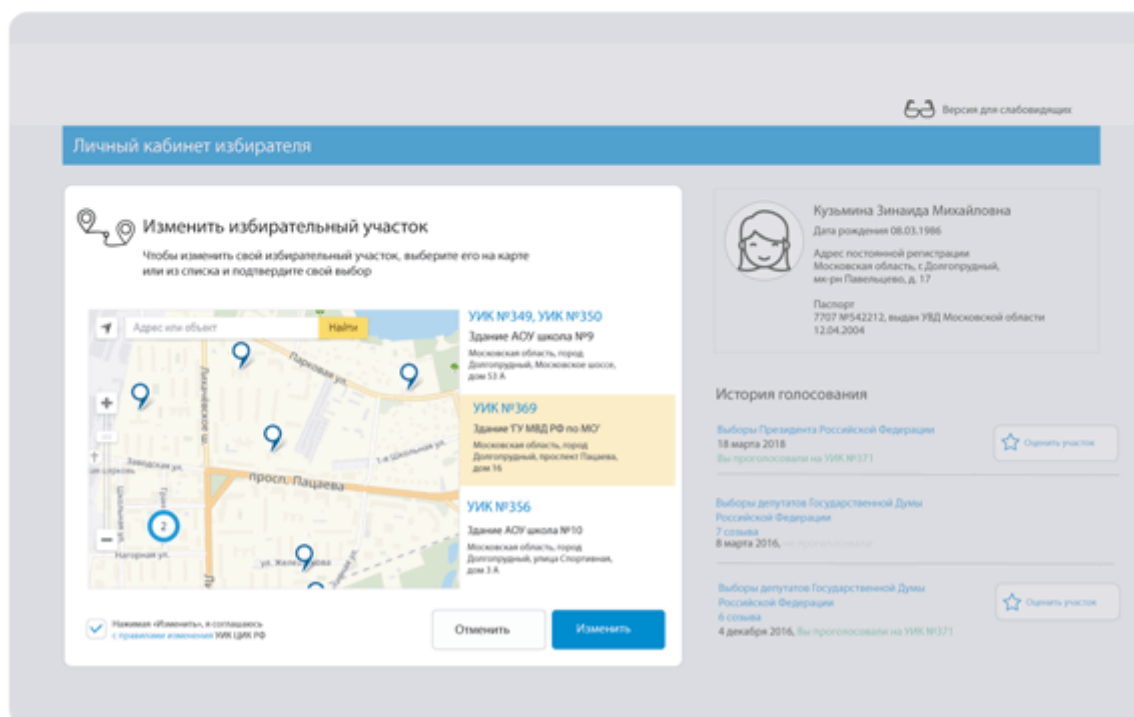


Рис. 1.12.

Личный кабинет избирателя (выбор участка)

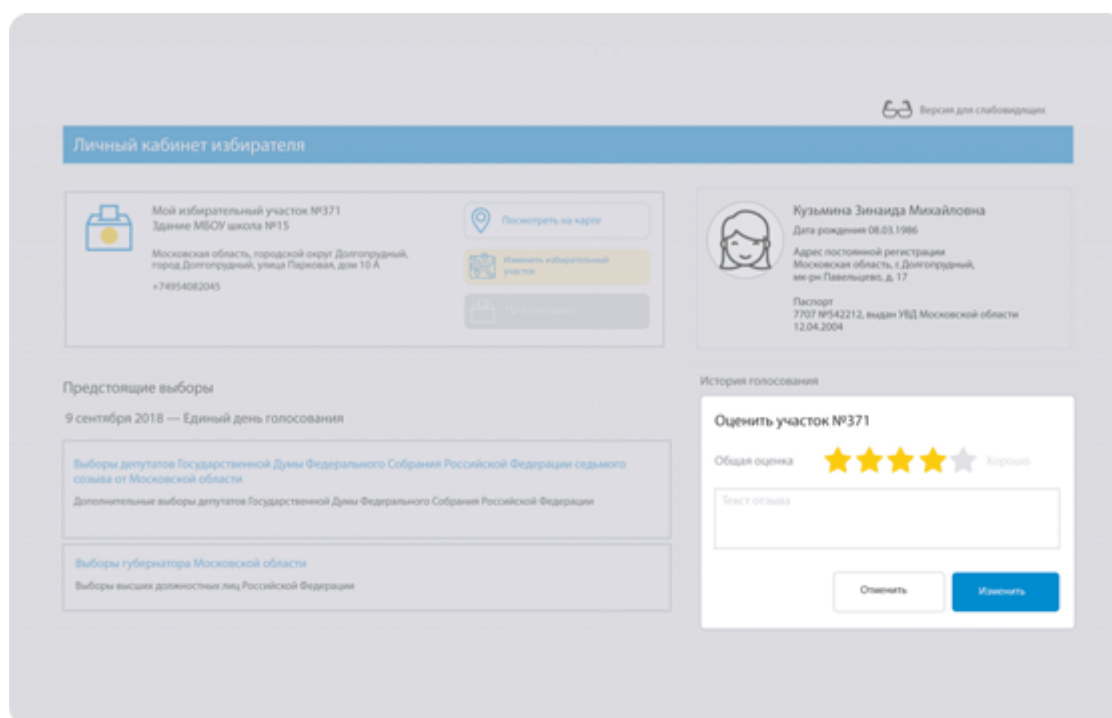


Рис. 1.13.
Личный кабинет избирателя (обратная связь)

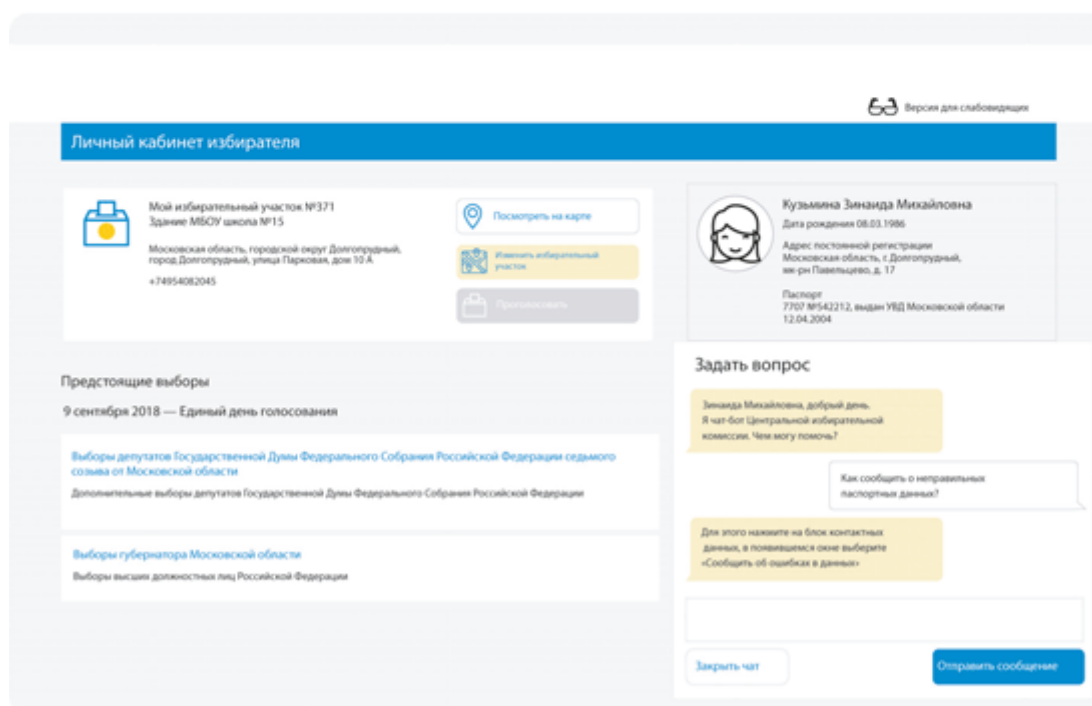


Рис. 1.14.
Личный кабинет избирателя (сервисы)

По нашему мнению, проект «Электронная трудовая книжка» – краеугольный камень в проекте «Цифровая экономика», поскольку отношения государства, работодателя и работника составляют основу экономических отношений во всем мире. В упрощенном виде эти отношения можно описать в виде схемы, приведенной на рисунке 1.15.

Подобный инструментариум предоставляет богатые возможности для взаимодействия всех участников

данного проекта. Для государства очевидны следующие:

- строгий учет и контроль предоставления социальных услуг в полном объеме;
- персональное предоставление льгот, адресная социальная поддержка;
- сквозной контроль социальных бюджетов всех уровней;
- возможность реализации реестровой модели государственных услуг;
- подтверждение доступа к персональным данным цифровой подписью гражданина: возможность сервисов «Единая медицинская карта», «Электронная трудовая книжка», различные интерпретации удостоверений льготников и т. д.;
- подтверждение работником факта смены работы – электронная идентификация и передача данных; онлайн-учет;
- реестр изменений и детальная аналитика по всем срезам данных держателя за любой период (как по работнику, группам работников, географии, так и по работодателю, отраслевому признаку и т. д.);
- аналитика социального состояния общества в реальном времени.



Рис. 1.15.

Электронная трудовая книжка

Ключевым фактором успеха подобного проекта, как было сказано ранее, служит инструментарий для работы с огромным массивом данных, который будет сформирован единовременно, а органам исполнительной власти на постоянной основе нужно будет работать с ним, поскольку по каждому объекту учета имеется огромное количество сценариев изменения данных и сами данные, которые необходимо поддерживать в актуальном состоянии (рис. 1.16).

Вот только краткое перечисление преимуществ, которые лежат на поверхности и позволяют решить огромное количество текущих проблем в данной области:

- электронная трудовая книжка не теряется;
- наличие всегда актуальных данных в различных интерпретациях личного кабинета;
- фиксируется все (образование, стаж, должность, договор, оклад, выплаты, начисления, курсы повышения квалификации и другая информация), что позволяет составить наиболее полную картину о работнике и работодателе (рис. 1.17).

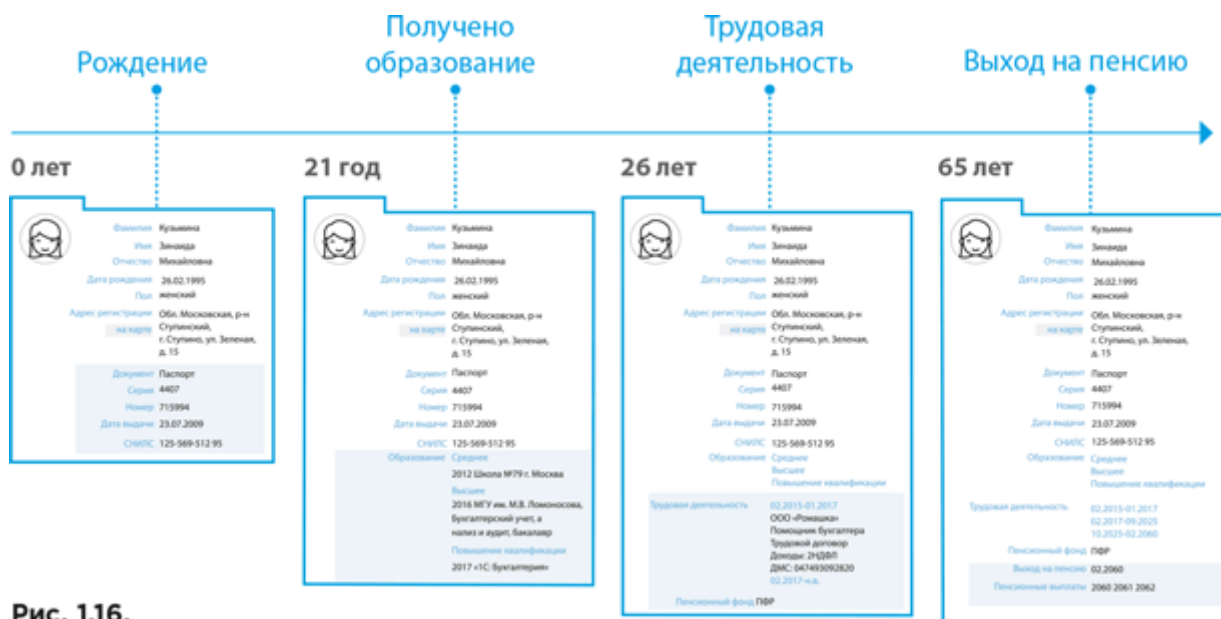


Рис. 1.16.

Данные о работнике

Также одна из ключевых задач создания электронной трудовой книжки – задача консолидации и гармонизации всего массива данных о работнике и работодателе в одной информационной системе, позволяющей обеспечить интероперабельность и непротиворечивость собранных данных из различных информационных систем целого ряда федеральных органов исполнительной власти (рис. 1.18).

Несмотря на организационную и техническую сложность проекта «Электронная трудовая книжка», его эффективность и возможность реализации может быть обеспечена за счет глубокого анализа текущих и проектирования будущих бизнес-процессов, профессиональной команды разработчиков, способных спланировать и реализовать архисложный процесс проектирования архитектуры, правильно сформулировать задачу и создать систему с огромным массивом данных.

Государству электронная трудовая книжка позволит сформировать массив данных для создания новых сервисов, которыми будут пользоваться федеральные и региональные органы исполнительной власти (рис. 1.19).

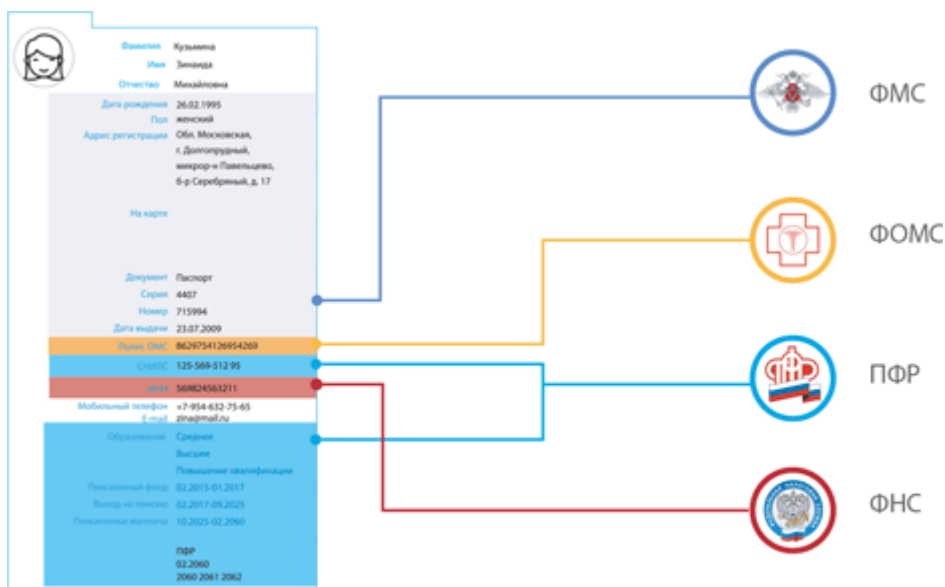
Отдельно стоит упомянуть новые аналитические возможности, предоставляемые электронной трудовой книжкой для формирования политики в области трудовых отношений, а также реализации краткосрочных, среднесрочных и долгосрочных планов в этой области (рис. 1.20).

Возможности информационной системы обеспечат также новым инструментарием коммерческие организации и позволят им эффективнее проводить свою политику в области работы с персоналом (рис. 1.21).

Одним из возможных инструментов могут быть средства поддержки мероприятий по поиску талантливых работников, специалистов редких областей и проведению конкурсов профессионального мастерства (рис. 1.22).

**Рис. 1.17.**

Фиксируемые сведения о работнике и работодателе

**Рис. 1.18.**

Консолидация и гармонизация данных

Нельзя не упомянуть, что электронная трудовая книжка – уникальный по своему удобству и эффективности инструмент для работников.

Мы неоднократно сталкивались с потребностью в подобных инструментах как для создания различных цифровых сервисов в целом, так и для реализации их в одной из самых чувствительных для граждан социальной сфере государственных услуг. Зачастую нехватка или отсутствие обобщенной информации о гражданах не позволяет государственным институтам своевременно и эффективно проводить мероприятия социальной политики. Только объективные критерии, основанные на данных, дают возможность в режиме реального времени видеть картину происходящих изменений в обществе, вовремя оказывать услуги, обеспечивать поддержку и социальную защиту, предлагать всевозможные льготы и субсидии гражданам, а также выявлять нарушения и отсутствие оснований для указанных мер.



Рис. 1.19.

Поддержка новых сервисов

С технологической точки зрения накопление в одном месте огромного массива данных о гражданах может быть одним из самых эффективных инструментов обеспечения их прав. Возможность ведения информации о гражданах с высоким и низким доходом, о недвижимости и транспортных средствах во владении, о составе семьи, а также многих других атрибутах позволяет рассматривать цифровизацию с позитивной и созидательной стороны. Повышение эффективности взаимодействия государства и общества позволяет сделать колоссальный шаг вперед, но для этого необходимо преодолеть опасения граждан в отношении «некорректных сценариев» использования персональных данных, обеспечив гарантии их неприкосновенности, согласование использования и утилизацию, поскольку только гражданин должен иметь право на использование кем бы то ни было своего «цифрового двойника» (рис. 1.23).

РЕГИОНЫ С НАИБОЛЬШИМ ОТТОКОМ СПЕЦИАЛИСТОВ С ВЫСШИМ ОБРАЗОВАНИЕМ

Аналитическая система → Движение трудовых ресурсов



Рис. 1.20.

Аналитические возможности

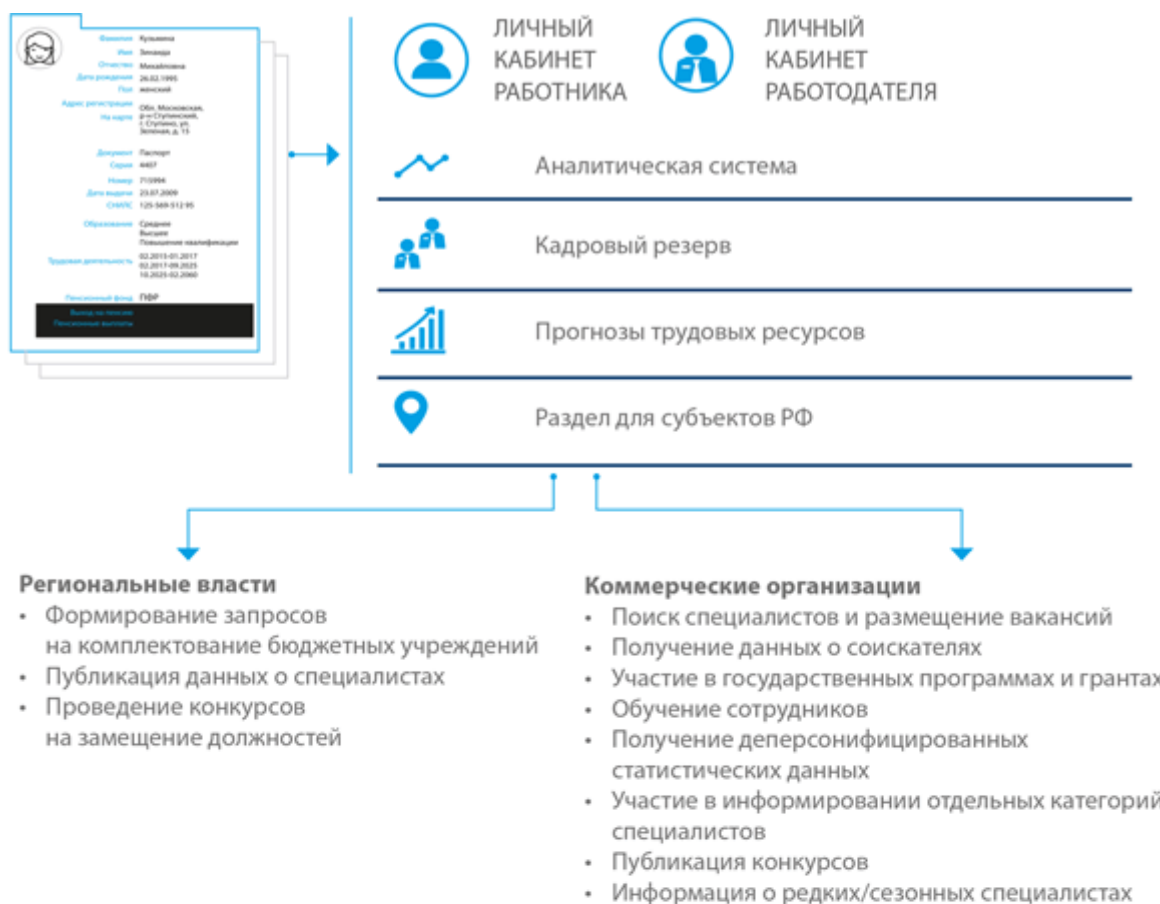


Рис. 1.21.

Новый инструментарий для коммерческих организаций

The form displays the following information for a specialist:

Личные данные:

- Фамилия: Новикова
- Имя: Зинаида
- Отчество: Михайловна
- Дата рождения: 26.02.1995
- Пол: женский
- Адрес регистрации: Обл. Московская, р-н Ступинский, г. Ступино, ул. Зеленая, д. 15
- Тип документа: Паспорт
- Серия: 4511
- Номер: 895001
- Дата выдачи: 02.10.2012

Образование: Высшее

Трудовая деятельность: 03.06.2011 — Бухгалтер; 05.07.2018 — Главный бухгалтер

Участие в государственных грантах: 2010 — «Молодые таланты 2010»

Рис. 1.22.

Поддержка поиска специалистов

Один из самых эффективных механизмов – создание инструментария работы с данными в рамках реализации сервисов, направленных как на существенное повышение эффективности деятельности государственного аппарата, так и на предоставление гражданам услуг в электронном виде. Это совершенно меняет модель взаимодействия государства и общества, а также позволяет экономить огромные средства и время аппарата государственного управления. Подобные инструменты позволяют сделать гигантский шаг

на пути избавления государства от различного рода ошибок и манипуляций с данными, на которых теряются огромные средства.

Для движения в этом направлении необходимо проанализировать текущее состояние данных, которые являются основными для администрирования со стороны государства. Объем накопленных в государственных информационных системах данных огромный, но работа с ними далека от идеала, например, из-за отсутствия документирования информационных систем (в части данных), использования устаревшей или неправильной архитектуры или закрытости систем. Эти проблемы присущи не только государственным, но и корпоративным информационным системам, архитекторы которых в погоне за сиюминутным результатом и персонифицированностью закладывают «мину замедленного действия» под возможностью развития систем и совместного использования данных для решения более широкого круга задач.

Одно из самых перспективных направлений работы с данными для государства и общества – создание информационных систем, актуальных практически для всех направлений деятельности. Сюда относятся такие решения, как «Цифровой профиль гражданина», «Цифровой профиль юридического лица», «Цифровой профиль изделия», «Цифровой профиль здания», «Цифровой профиль транспортного средства». Подобный подход применим фактически для всех объектов со сложной описательной частью.

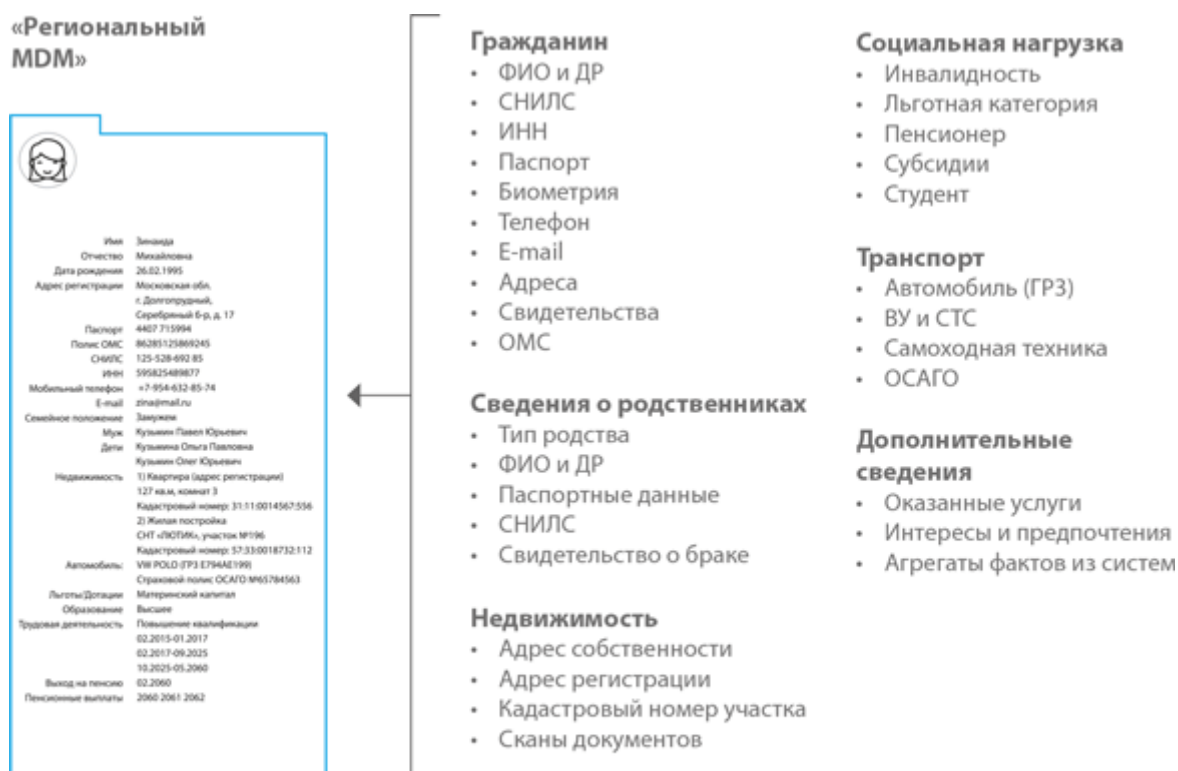


Рис. 1.23.

Обрабатываемые данные о гражданине

Уже сейчас проекты цифровизации приносят ощутимую пользу, и в ближайшем будущем именно эти направления будут одними из наиболее востребованных. В качестве примера можно привести преимущества проекта «Цифровой профиль гражданина».

1. Единая точка концентрации данных.

1.1. Сведения о гражданине – цифровой профиль.

1.2. Сведения об объектах, относящихся к гражданину (движимое и недвижимое имущество).

1.3. Права и обязанности граждан.

1.4. Действия граждан (использование и оплата различных государственных сервисов: налоги, субсидии, штрафы и т. п.).

2. Обработка и аналитика.

- 2.1. Система помощи принятия и поддержки управленческих решений.
- 2.2. Подготовка прогнозов на основе данных.
- 3. Контроль качества данных.
- 3.1. Единые стандарты хранения и обработки данных.
- 3.2. Контроль соблюдения единых стандартов данных.

В качестве ключевых целей внедрения решения «Цифровой профиль гражданина» можно выделить следующие:

- сокращение расходов бюджета – путем обработки и аналитики данных выявление неэффективного расходования средств;
- аргументированное принятие решений – повышение качества данных, на основе которых принимаются ключевые решения.

При объединении сведений о гражданах из различных источников государственные органы неминуемо столкнутся с конфликтами в данных, что позволит выявить либо недобросовестных граждан, получающих поддержку государства, которая им не положена, либо граждан, имеющих право на господдержку, но по каким-то причинам ее не получающим.

Также «Цифровой профиль гражданина» обеспечивает взаимодействие граждан и государства в режиме онлайн, предоставляя предиктивные услуги и всевозможные сервисы по информированию.

Отдельно стоит отметить использование данных в спорте – особой сфере деятельности как общества, так и государства, которое обеспечивает и формирует среду для активного образа жизни.

Управление спортом, если его рассматривать как информационную систему, имеет много общего с аналогичными системами из других областей, но здесь есть и свои отраслевые отличия. Если говорить о спорте, то наверняка каждый читатель сразу назовет большое количество категорий данных: это различные спортивные дисциплины, федерации, тренеры, спортсмены, медицинский персонал, спортивные объекты и сооружения, разряды, названия спортивных соревнований и мероприятий, результаты, научные и методические работы. Каждая категория включает в себя сотни атрибутов и тысячи показателей. Многие из них взаимосвязаны, что позволяет на основе данных и выявленных в ходе их анализа закономерностей планировать и реализовывать политику в области управления спортом.

Литература к главе 1

- DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)
- Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)
- Stephens-Davidowitz, S. Everybody lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. 2017.

Глава 2. От данных к мудрости

2.1. Данные и информация

Прежде чем начать разбираться в вопросах управления и извлечения ценности из данных, стоит подробнее остановиться на терминах «данные» (data) и «информация» (information) и их значениях. Применительно к сфере управления данными эти термины часто используют как синонимы, но во многих

случаях значения различают²⁷²⁸. Понять разницу поможет экскурс в историю их происхождения.

В английском языке слово *information* появилось гораздо раньше, чем слово *data*, – в конце XIV века. Слово пришло из старофранцузского языка. Французские *information*, *informateur* – производные от *informer* – «осведомлять, сообщать, делать известным». Первоисточник – латинское *informatio* – «разъяснение, истолкование, сообщение». От *in* – «в, к, в направлении» + *forma* – «форма». Истоки слова «форма» неизвестны²⁹³⁰.

Во второй половине XV века за словом *information* закрепилось значение «переданные сведения, относящиеся к определенной теме». Сфера применения постепенно расширялась. В частности, в XX веке слово стало появляться в документальных источниках в связи с передачей телевизионных сигналов – с 1937 года, применением перфокарт – с 1944-го, анализом ДНК – с 1953-го. Термин «теория информации» (*information theory*) используется с 1950 года, «информационная технология» (*information technology*) – с 1958-го (впервые появился в журнале *Harvard Business Review*), «информационная революция» (*information revolution*) – с 1966-го, «информационная перегрузка» (*information overload*) – с 1967-го.

В русский язык слово «информация» пришло из западноевропейских языков при Петре I – в начале XVIII века (но в словарях появляется лишь во второй половине XIX века). Основные значения – «сообщения, сведения, пояснения».

Что же касается слова *data*, то оно в английском языке стало использоваться (в значении, близком к сегодняшнему) в 1640-х годах – более чем на два столетия позже слова *information*.

Слово *data* произошло от латинского *datum* – «данная вещь» (от латинского *dare* – «давать»). Оно применяется как во множественном, так и в единственном числе, однако в научной литературе для единственного числа может использоваться термин *datum*.

1640-е годы – время начала Английской гражданской войны, открывшей путь к промышленной революции XVIII–XIX веков. Одновременно это было время осуществления первой научной революции и возникновения классической европейской науки. Слова *data* и *datum* в значении «данные или предоставленные факты» впервые появились в работах английского священника Генри Хаммонда (Henry Hammond), который, как и многие представители его профессии в то время, участвовал в формировании основ классической науки³¹³².

Начиная с работ Хаммонда слово *information* стало все чаще и чаще дублироваться словом *data*, чему способствовала научная революция, наступление «Века разума» (*The Age of Reason*) и эпохи Просвещения. Появилось (благодаря быстрому распространению книгопечатания, изобретенного Гуттенбергом в 1450 году) большое количество книг по уже существовавшим и вновь появляющимся отраслям науки. Таким образом, был создан и активно обрабатывался огромный объем совместно используемых сведений. Для обозначения таких сведений, предоставляемых для обсуждения и осмысления, начал применяться термин *data* (сначала в классическом значении – «факты, данные в качестве основы для вычислений при решении математических задач»), более подходящий в данном случае, чем термин *information*³³.

27 DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

28 Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Д. Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

29 <https://www.etymonline.com/word/information>.

30 <https://ecanet.ru/word/Информация>.

31 <https://circa-navigate.corsairs.network/a-daring-history-of-data-4ae39ba9459>.

32 В частности, Хаммонд перевел на английский язык ряд сочинений французского ученого и философа Блеза Паскаля.

33 <https://circa-navigate.corsairs.network/the-history-of-data-5663fc4f1398>.

Необходимо отметить еще одно важное слово, появление которого оказало существенное влияние на дальнейшее распространение термина «данные». В 1798 году Джон Синклер³⁴ в своем «Статистическом отчете о Шотландии» ввел в английский язык термин «статистика». По определению Синклера, статистика – это собрание фактов, которые совершенно не обязательно должны были быть представлены (и даже в основном не были представлены) в числовом виде. Однако с 1829 года слово statistics стало употребляться в английских документальных источниках в более строгом значении – «собранные и классифицированные числовые данные». С этого момента изучение любого предмета стало сопровождаться проведением всеобъемлющих подсчетов³⁵.

Широкое распространение статистических методов привело к тесной смысловой привязке термина data к термину statistics. В 1897 году в документальных источниках на английском языке слово data стало использоваться в новом значении – «факты в числовом виде, собранные для дальнейшего использования».

В XX веке спектр значений слова data существенно расширился.

В 1946 году появилось значение «передаваемая и хранимая информация, с помощью которой осуществляется выполнение компьютерных операций»³⁶.

Стали использоваться новые термины: «обработка данных» (data processing) – с 1954 года, «база данных» (database) – с 1962 года, «ввод данных» (data entry) – с 1970 года³⁷.

В русском языке слово «данные» – производное от «дать», «давать». Родственно древнеиндийскому слову datis – «дар». Данные – то, что дано, найдено. В толковом словаре Даля «данныя, данности – известное, бесспорное, очевидное, верное, все, что служит основанием для какого-либо вывода, расчета, заключения»³⁸. В энциклопедическом словаре Брокгауза и Ефрона приведено следующее определение: «В вопросах математики данные суть величины, значения которых известны или предполагаются известными; зная их, требуется в рассматриваемом вопросе определить искомые неизвестные величины».

Данные (Δεδομένα) есть заглавие одного из сочинений Эвклида, составляющего продолжение его «Элементов». Можно указать несколько изданий этой книги: Гарди в 1625 году с греческим текстом и Баррова в 1659 году Эвклид называет данным все то, что на основании теорем, заключающихся в элементах, непосредственно следует из условий задачи. Например, если проводим из данной точки прямую, касательную к данному кругу, то эта прямая есть данная по величине и положению»³⁹.

Приведем несколько современных определений.

Согласно «Оксфордскому словарю современного английского языка»⁴⁰:

- Data: 1) Известные факты, используемые для вывода или расчета. 2) Числовые и нечисловые значения характеристик кого-либо (чего-либо), с которыми выполняет операции компьютер или какое-нибудь другое подобное устройство.

- Information: 1) а) Что-то, что было сообщено; знания. б) Элементы знаний; новости. 2) Обвинение или жалоба, поданная в суд и т. п.

34 Сэр Джон Синклер (Sir John Sinclair) – шотландский политический деятель и экономист, первый президент Совета сельского хозяйства Шотландии.

35 <https://www.etymonline.com/word/statistics>.

36 В это время в США проводились работы по созданию первого в мире лампового компьютера, в ходе которых была предложена и реализована знаменитая «архитектура фон Неймана», основанная на принципе совместного хранения команд и данных в компьютерной памяти и применяемая сегодня в подавляющем большинстве вычислительных машин.

37 <https://www.etymonline.com/word/data>.

38 <https://ecanet.ru/word/Данные>.

39 http://1slovar.ru/d/brokgauz_efron/122384/.

40 Oxford Dictionary of Current English / Edited by Della Thompson: 2nd Edition, Oxford University Press, 1993.

Согласно «Новому словарю русского языка»⁴¹:

• Данные: 1) Сведения, факты, характеризующие кого-либо, что-либо, необходимые для каких-либо выводов, решений. 2) Свойства, способности, качества как условия или основания, необходимые для чего-либо.

• Информация: 1) Сообщение о положении дел где-либо, о состоянии чего-либо. 2) а) Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальными устройствами. б) Обмен такими сведениями между людьми и специальными устройствами. в) Обмен сигналами в животном и растительном мире. 3) То же, что: информирование.

Все это позволяет понять разницу между понятиями «данные» и «информация» и очертить преимущественные сферы их применения.

Следует заметить, что здесь в отношении термина «информация» мы ограничиваемся его общепринятыми нестрогими определениями и не обсуждаем научные определения, используемые в теории связи, теории информации и кибернетике и связанные с понятием количества информации. Отметим только, что известны два основных сугубо научных подхода к определению этого термина⁴².

Согласно первому, информация – это внешнее описание объекта, содержащееся в передаваемом сообщении. Количество информации определяется как мера неопределенности, снимаемой полученным сообщением. Такое представление в 1948 году ввел Клод Шеннон⁴³ в своей статье «Математическая теория связи», попутно предложив для обозначения наименьшей единицы информации использовать слово «бит».

Согласно второму подходу, информация не связывается с передаваемым сообщением. Она относится к объекту и есть мера его разнообразия. Автор этого представления – Уильям Эшби⁴⁴.

Указанные научные определения в дальнейшем в этой книге не используются, поскольку ценность информации слабо связана с ее количеством. Важность информации субъективна и зависит от цели ее потребителя. Все способы количественного определения ценности основаны на представлении о цели, достижению которой способствует полученная информация. Чем больше информация помогает достижению цели, тем более ценной она считается⁴⁵. Массив информации объемом в сотни гигабайт может иметь колоссальную важность для одной организации и нулевую для другой.

Основываясь на приведенных сведениях, можно продолжить обсуждение вопросов, относящихся к ценности данных.

Как уже говорилось выше, применительно к сфере управления данными термины «данные» и «информация» часто используются как синонимы. Кроме того, как видно из приведенных выше определений, термин «информация» иногда заменяется термином «знания».

Всегда ли допустимо такое обобщение и не приводит ли оно к недоразумениям? В частности, к существенному недопониманию между участниками многочисленных проектов по внедрению различного рода информационных систем, а также между участниками их дальнейшей эксплуатации и пользователями. В том числе и по вопросам определения ценности данных.

Для ответа необходимо более подробно обсудить взаимосвязь понятий «данные» и «информация», а также соотнести их с понятиями «знания» (knowledge) и «мудрость» (wisdom).

2.2. Информационная иерархия

41 Ефремова Т. Ф. Новый словарь русского языка. Толково-словообразовательный. В 2 т. – М.: Русский язык, 2000.

42 Шамис А. Л. Модели поведения, восприятия и мышления: курс лекций. – М.: Интуит НОУ, 2016. – URL: <https://intuit.ru/studies/courses/2191/423/info>.

43 Клод Шеннон (Claude Elwood Shannon) – выдающийся американский математик, основатель математической теории информации.

44 Уильям Эшби (William Ross Ashby) – английский психиатр, крупный теоретик кибернетики.

45 Чернавский Д. С. Синергетика и информация (динамическая теория информации). 2-е изд., испр. и доп. – М.: Едиториал УРСС, 2004.

При обсуждении взаимосвязи понятий «данные», «информация», «знания» и «мудрость» обычно используется классическое представление – иерархия Data-Information-Knowledge-Wisdom (DIKW). Часто в литературе применяются названия «иерархия знаний», «пирамида знаний» или же «информационная иерархия», «информационная пирамида».

В большинстве исследований, посвященных происхождению иерархии DIKW (например,⁴⁶⁴⁷⁴⁸) в качестве одного из наиболее ранних источников, в которых встречается ее описание, обычно упоминается статья Харлана Кливленда⁴⁹ «Информация как ресурс». Она была опубликована в 1982 году в американском журнале *The Futurist*⁵⁰.

По мнению Кливленда, данные образуются в результате исследований, а также путем создания или сбора или же совершения открытий, в то время как информация имеет контекст. Данные превращаются в информацию с помощью организации их таким образом, чтобы мы могли легко делать какие-либо заключения. Кроме того, они преобразуются в информацию с помощью представления (презентации), например в визуальной или аудиальной (звуковой) форме.

Хотя известно несколько более ранних публикаций других авторов, затрагивающих эту тему⁵¹, статья Кливленда цитируется особенно часто и примечательна фрагментом из поэмы Томаса Элиота⁵² «Камень» (The Rock), написанной в 1934 году, в которой отражена иерархическая взаимосвязь понятий «информация», «знания» и «мудрость».

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
Где Жизнь, которую мы потеряли в жизни?
Где мудрость, которую, мы потеряли в знаниях?
Где знания, которые мы потеряли в информации?

Кливленд назвал определенную в поэме цепочку понятий «иерархией Элиота». После выхода его статьи в специальной литературе прочно закрепилось мнение о том, что именно Элиоту принадлежит авторство идеи иерархии DIKW (или, по крайней мере, IKW)⁵³.

46 Sharma N. The origin of the data information knowledge wisdom hierarchy, 2005. – URL: https://web.archive.org/web/20051221151426/http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm.

47 Rowley J. The wisdom hierarchy: representation of the DIKW hierarchy. *Journal of Information Science*, 2007, 33(2), 163–180. DOI: 10.1177/0165551506070706. – URL: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>.

48 Williams D. Models, Metaphors and Symbols for Information and Knowledge Systems // *Journal of Entrepreneurship Management and Innovation (JEMI)*, 2014, Vol. 10, Iss. 1: 79–107. DOI: 10.7341/20141013. – URL: <https://www.semanticscholar.org/paper/Models-%2C-Metaphors-and-Symbols-for-Information-and-Williams/fde0d2b7da419d66f1cdbc7cce8ff37116254cf3>.

49 Харлан Кливленд (Harlan Cleveland) – американский дипломат, педагог и писатель.

50 Cleveland H. Information as Resource. *The Futurist*, 1982. Vol. 16. No. 6, Pp. 34–39.

51 Williams D. Models, Metaphors and Symbols for Information and Knowledge Systems // *Journal of Entrepreneurship Management and Innovation (JEMI)*, 2014, Vol. 10, Iss. 1: 79–107. DOI: 10.7341/20141013. – URL: <https://www.semanticscholar.org/paper/Models-%2C-Metaphors-and-Symbols-for-Information-and-Williams/fde0d2b7da419d66f1cdbc7cce8ff37116254cf3>.

52 Томас Элиот (Thomas Stearns Eliot) – выдающийся английский поэт, лауреат Нобелевской премии по литературе 1948 года.

53 В литературе часто упоминается еще один представитель мира искусств, который описал иерархию IKW до того, как она стала активно обсуждаться специалистами по обработке информации и управлению знаниями, – легендарный американский рок-музыкант Фрэнк Заппа (Frank Zappa). Связанные с этой темой рассуждения приводятся в его песне Packard Goose из альбома Joe's Garage: Act II & amp; III (1979).

Среди академических работ, в которых рассматривается иерархия DIKW, наиболее ранней стала статья Милана Желены⁵⁴ «Системы поддержки управления: На пути к интегрированному управлению знаниями»⁵⁵. В ней описана схема последовательного продвижения от данных к знаниям. Ученый предложил упрощенные определения уровней иерархии, основанные на знании ответов на простые вопросы.

- Данные – «не знаю – ничего» (know-nothing).
- Информация – «знаю – что» (know-what).
- Знания – «знаю – как» (know-how).
- Мудрость – «знаю – почему» (know-why).

В то же время Желены заметил, что, хотя данные и информация (благодаря их атомарной, дробной, нецелостной природе) могут быть сгенерированы без интерпретации человеком, знания и мудрость зависят от человека и контекста и не могут быть рассмотрены без использования процедур оценки и принятия решения человеком.

Автором одного из наиболее систематизированных и целостных описаний взаимосвязи понятий «данные», «информация», «знания» и «мудрость» часто называют Рассела Акоффа⁵⁶. В 1989 году был опубликован текст его президентского обращения к Международному обществу общесистемных исследований (International Society for General Systems Research, ISGSR) под заголовком «От данных к мудрости»⁵⁷.

Хотя размышления Акоффа относятся к иерархии типов содержания человеческого разума, они справедливы и по отношению к информационным системам. Иерархия описывается следующим образом. «Мудрость» находится на вершине иерархии. Ниже следуют «понимание»⁵⁸, «знание», «информация» и в самом низу – «данные». Каждый из уровней включает содержание уровней, расположенных ниже, – например, не может быть мудрости без понимания и понимания без знания.

Акофф предлагает следующие определения данных, информации, знаний и мудрости, а также связанных с ними процессов трансформации.

- Данные – это символы, представляющие свойства объектов, событий и их окружения. Это результаты наблюдений. Наблюдать означает ощущать. Технология ощущений, ее инструментализация, конечно, высоко развита у человека от природы. Информация, как уже отмечалось, извлекается из данных с помощью их анализа, во многих аспектах которого компьютеры превосходны.

Данные, подобно железной руде, не имеют ценности, пока они не преобразуются в соответствующую полезную форму. Поэтому разница между данными и информацией не структуральная, а функциональная, данные обычно редуцируются при их преобразовании в информацию.

- Информация состоит из описаний, ответов на вопросы, начинающиеся с таких слов, как «кто», «что», «где», «когда», «сколько». Информационные системы генерируют, запоминают, извлекают и обрабатывают данные. Во многих случаях обработка носит статистический или арифметический характер.

⁵⁴ Милан Желены (Milan Zeleny) – американский экономист чешского происхождения, профессор в области систем управления Фордхемского университета (Fordham University).

⁵⁵ Zeleny M. Management support systems: Towards integrated knowledge management. Human Systems Management, 1987, 7(1), 59–70.

⁵⁶ Рассел Акофф (Russell Lincoln Ackoff) – известный американский ученый и консультант в области исследования операций, теории систем и стратегического менеджмента.

⁵⁷ Ackoff R. L. From Data to Wisdom // Journal of Applied Systems Analysis, 1989, 16, 39. (Русский перевод: Р. Акофф. От данных к мудрости // Проблемы управления в социальных системах, 2011. – URL: <https://cyberleninka.ru/article/n/ot-dannyh-k-mudrosti>)

⁵⁸ В иерархию DIKW Р. Акофф добавил еще один элемент – «понимание» (understanding). Он поместил его между «знаниями» и «мудростью» (т. е. получилась иерархия DIKUW). Согласно Акоффу, понимание предполагает вынесение диагноза и предписаний (взаимодействие с физическим миром с помощью навыков и знаний).

В любом случае информация выводится из данных.

- Знание есть ноу-хау, например о том, как система работает. Знание позволяет преобразовать информацию в инструкции. Оно делает возможным контроль над системой.

Знание может быть обретоено двумя путями: либо получением его от тех, кто им обладает, либо извлечением из собственного практического опыта. В любом случае приобретение знания есть обучение. Когда программируются компьютеры, они «обучаются», как сделать что-то.

- Мудрость есть способность увеличивать эффективность по цели, (effectiveness), в то время как информация и знание сосредоточены на эффективности по средствам (efficiency)⁵⁹.

Мудрость имеет большую ценность благодаря ментальной функции, которую мы называем суждением. Все оценки эффективности основаны на логике, которая может быть определена, а значит, запрограммирована и автоматизирована. Эти принципы общи и объективны. Мы можем говорить об эффективности действия безотносительно к исполнителю. А по отношению к суждению это не так. Ценность действия всегда зависит от того, кто действует, редко когда она одинакова для двух исполнителей, даже если они делают одно и то же. Эффективность по средствам не связана с мотивами деятельности; а эффективность по целям определяется именно этическими и эстетическими ценностями. Они являются уникальными и персональными.

По мнению Акоффа, элементы иерархии DIKW имеют временное измерение. Информация, подобно новостям, довольно быстро стареет. Знание живет дольше, хотя и оно неизбежно устаревает. Мудрость имеет вечную значимость для человечества (если она не утрачивается).

Позже американские эксперты в области теории организаций (Bellinger и др.⁶⁰) выступили с критикой расширения Акоффом иерархии DIKW за счет «понимания», уточнив, что «понимание» не является отдельным уровнем, а скорее поддерживает переходы с предыдущих уровней на следующие (рис. 2.1).

⁵⁹ В тексте русского перевода статьи Акоффа есть следующее примечание переводчика: «В английском языке разными словами обозначены “эффективность по достижению промежуточных целей, т. е. в использовании средств достижения конечной цели” – efficiency и “эффективность в достижении самой конечной цели” – effectiveness. В дальнейшем будем переводить эти термины как “эффективность по средствам” и “эффективность по цели”. К этому замечанию можно еще добавить, что efficiency часто переводится как «производительность», а effectiveness – как «действенность» *.

* Ackoff R. L. From Data to Wisdom // Journal of Applied Systems Analysis, 1989, 16, 39. (Русский перевод: Р. Акофф. От данных к мудрости // Проблемы управления в социальных системах, 2011. – URL: <https://cyberleninka.ru/article/n/ot-dannyh-k-mudrosti>)

⁶⁰ Bellinger G., Castro D., Mills A. Data, Information, Knowledge, & Wisdom, 2004. – URL: <http://www.systems-thinking.org/dikw/dikw.htm>.



Рис. 2.1.

Роль понимания в обеспечении продвижения по иерархии DIKW*

* Bellinger G., Castro D., Mills A. Data, Information, Knowledge, & Wisdom, 2004. – URL: <http://www.systems-thinking.org/dikw/dikw.htm>.

Согласно приводимому исследованию данные представляют факты или утверждения о событии без связи с другими вещами.

Пример. Идет дождь.

Информация содержит понимание какой-либо связи, возможно причинно-следственной.

Пример. Температура упала на 15 градусов, а потом пошел дождь.

Знания отражают паттерн, который связывает факты и обычно обеспечивает высокий уровень предсказуемости относительно того, что описано или что произойдет дальше.

Пример. Если влажность очень высокая и температура существенно падает, то влага вряд ли сможет удержаться в атмосфере, поэтому обычно идет дождь.

Мудрость воплощает более глубокое понимание фундаментальных принципов, содержащихся в знании, которые, по сути, являются основой знаний. Мудрость носит системный характер.

Пример. Идет дождь, потому что идет дождь. И это включает в себя понимание всех взаимодействий, которые происходят между дождем, испарением, воздушными потоками, температурными градиентами и сопровождающими их изменениями.

Чаще всего иерархия DIKW изображается в виде пирамиды (рис. 2.2). Такое графическое представление впервые появилось в работе⁶¹ (об этом, в частности, сообщается в обзорной статье⁶²).

⁶¹ Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link, published at Intergovernmental Oceanographic Commission (UNESCO) – OceanTeacher: a training system for ocean data and information management, 2004. – URL: https://web.archive.org/web/20071202033948/http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DIKWchain.pdf.

⁶² Williams D. Models, Metaphors and Symbols for Information and Knowledge Systems // Journal of Entrepreneurship Management and Innovation (JEMI), 2014, Vol. 10, Iss. 1: 79–107. DOI: 10.7341/20141013. – URL: <https://www.semanticscholar.org/paper/Models-%2C-Metaphors-and-Symbols-for-Information-and-Williams/fde0d2b7da419d66f1cdbc7cce8ff37116254cf3>.



Рис. 2.2.

Пирамида знаний*

* Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link, published at Intergovernmental Oceanographic Commission (UNESCO) – OceanTeacher: a training system for ocean data and information management, 2004. – URL: https://web.archive.org/web/20071202033948/http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DIKWchain.pdf.

2.3. Внутри пирамиды знаний

Чтобы составить достаточно полную картину взаимоотношений между уровнями пирамиды знаний в соответствии с современными общепринятыми взглядами, приведем более подробное описание этих уровней. Для этого воспользуемся сведениями из часто цитируемых обзоров наиболее известных монографий и руководств по информационным системам и управлению знаниями⁶³⁶⁴⁶⁵⁶⁶.

Начнем с определений.

1. Определение данных.

Чаще всего в определениях данных отмечается следующее:

- Данные – это дискретные, объективные факты или наблюдения, неорганизованные и необработанные, не передающие никакого конкретного смысла.
- Элементы данных – это простейшие записанные фрагменты описаний вещей, событий, действий и транзакций.

Данные не имеют смысла или ценности, потому что они лишены контекста и интерпретации.

Интересно, что выделяется преимущественно то, чего не хватает данным, – они лишены смысла или

⁶³ Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link, published at Intergovernmental Oceanographic Commission (UNESCO) – OceanTeacher: a training system for ocean data and information management, 2004. – URL: https://web.archive.org/web/20071202033948/http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DIKWchain.pdf.

⁶⁴ Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 2007, 33(2), 163–180. DOI: 10.1177/0165551506070706. – URL: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>.

⁶⁵ Williams D. Models, Metaphors and Symbols for Information and Knowledge Systems // Journal of Entrepreneurship Management and Innovation (JEMI), 2014, Vol. 10, Iss. 1: 79–107. DOI: 10.7341/20141013. – URL: <https://www.semanticscholar.org/paper/Models-%2C-Metaphors-and-Symbols-for-Information-and-Williams/fde0d2b7da419d66f1cdbc7cce8ff37116254cf3>.

⁶⁶ Van Meter, Heather J. Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom. Law Technology and Humans, 2020, Vol. 2. No. 2, 69–80. DOI: 10.5204/lthj.1470. – URL: <https://lthj.qut.edu.au/article/view/1470>.

ценности, не организованы и не обработаны. Ввиду этих особенностей определения информации формулируются в терминах данных.

2. Определение информации.

Определения информации, как правило, опираются на ее взаимосвязь с данными. При этом используются такие понятия, как «формат», «структура», «организация», «смысл» и «ценность».

● Информация – это форматированные данные, которые могут быть определены как представление реальности.

● Информация – это данные, которые улучшают понимание предмета.

● Информация – это данные, представленные в форме, позволяющей им быть полезными и значимыми для людей.

● Информация – это данные, обработанные с определенной целью.

● Информация – это данные, которым придан смысл посредством добавления контекста.

Таким образом, информация в большинстве случаев определяется в терминах данных и рассматривается как организованные или структурированные данные⁶⁷. Их обработка обеспечивает соответствие данных конкретной цели или контексту и тем самым делает их значимыми, ценными, полезными и релевантными.

3. Определение знаний.

Определения знаний обычно сложнее, чем определения данных или информации, поскольку, как правило, в литературе об управлении знаниями много внимания уделяется расширенному обсуждению природы знаний, а также различным вариантам их представления и проявления. Это делает отбор отличительных свойств знаний труднее, чем свойств данных и информации. В частности, встречаются такие высказывания:

● «Знания – по своей сути неясный и неоднозначный термин».

● «До сих пор нет единого мнения о природе знаний, за исключением того, что они основаны на восприятии, которое может дать им рациональное обоснование».

Тем не менее знания часто определяются в терминах данных и информации.

● «Знания – это совокупность данных и информации, к которым добавляются экспертные мнения, навыки и опыт, в результате чего получается ценный актив, который может быть использован для содействия принятию решений».

● «Знания – это данные и/или информация, которые были организованы и обработаны для передачи понимания, опыта, накопленных результатов обучения и компетенции, так чтобы их можно было применить при решении текущих проблем или в текущей деятельности».

● «Знания основываются на информации, извлеченной из данных. В то время как данные отражают свойства вещей, знания относятся к способностям людей, предрасполагающим их действовать определенным образом».

С точки зрения эпистемологии разделяются неявные знания и явные⁶⁸.

К неявным (неформализованным) знаниям (*tacit knowledge*) относятся опыт, мастерство, культура мышления, интуиция, хранящиеся в нейронных структурах головного мозга как результат генетической наследственности, образования и полученных жизненных уроков. Неявные знания существуют в умах специалистов, развиваясь во времени, через опыт, почерпнутый из профессиональной деятельности, книг, наставничества, а также обучения. Эти знания зависят от жизненных ресурсов личности, от ее биофизических свойств и психологического потенциала. Они заложены в индивиде и не могут быть легко переданы другим.

К явным (формализованным) знаниям (*explicit knowledge*) относятся описания теорий, методов, методик, технологий, механизмов и машин, конструкций, систем и т. п. Эти знания хранятся на реальных физических носителях. Они представлены в виде книг, бумажных документов, рисунков, схем, фильмов,

67 В этой связи часто обращают внимание на то, что первоисточник слова *information* – латинское *informatio* – разъяснение, истолкование, сообщение. От *in* – «в, к, в направлении» + *forma* – «форма» (см. раздел 2.1. «Данные и информация»).

68 Нонака И., Такеучи Х. Компания – создатель знания. Зарождение и развитие инноваций в японских фирмах. – М.: Олимп-Бизнес, 2011.

аудио– и видеозаписей, электронных файлов, баз данных и т. п.

Знания существуют в континууме между неявными и явными знаниями. Явные знания – это неявные знания, которые были задокументированы⁶⁹.

Если суммировать встречающиеся определения, то знания можно рассматривать как совокупность данных, информации, понимания, умений, опыта, навыков и ценностей. Важно отметить, что не все авторы упоминают все эти элементы.

4. Определение мудрости.

К слову сказать, далеко не все авторы, описывающие иерархию DIKW, предлагают определение мудрости.

- Мудрость – это накопленные знания, позволяющие понять, как применять идеи и представления из одной области к новым ситуациям или проблемам.

- Мудрость – это высший уровень абстракции вкупе с предвидением и способностью смотреть за горизонт.

- Мудрость – это способность действовать, подходя критически или практически в любой конкретной ситуации. Она основана на этических суждениях, связанных с системой убеждений индивида.

Очевидно, что мудрость – еще более неуловимое и расплывчатое понятие, чем знания. Оно гораздо сильнее связано с человеческой интуицией, пониманием, интерпретацией и действиями, чем с информационными системами. При этом понятия «данные», «информация» и «знания» в определениях мудрости могут и не использоваться.

Наше обсуждение определений из различных источников показывает наличие консенсуса в отношении того, что по крайней мере данные, информация и знания должны определяться в терминах друг друга, хотя в качестве входных материалов для знаний могут выступать и данные, и информация.

Рассмотрев определения элементов иерархии DIKW, коротко обсудим процессы преобразования одного элемента в другой.

1. Преобразование данных в информацию.

Как уже отмечалось, согласно общепринятым представлениям, информация рассматривается как организованные или структурированные данные. Такая обработка обеспечивает соответствие данных конкретной цели или контексту и тем самым делает их значимыми, ценными, полезными и релевантными. Другими словами, структурирование данных в соответствии с некоторой схемой, имеющей значение и актуальность для отдельного человека, сообщества или задачи, наделяет их смыслом или, возможно, потенциалом для осмысления.

Важно заметить, что все данные (как только они собраны или переданы на хранение) в информационных системах и нашем сознании наделяются определенной структурой. Информационные системы всегда кодифицируют любой элемент данных для того, чтобы поместить его в базу данных, а затем найти для последующего использования. Люди, собирающие данные, должны разобраться в них, чтобы обеспечить их хранение с учетом взаимосвязей с уже накопленной информацией. Таким образом, если информация от данных отличается наличием структуры, то получается, что мы храним как в нашем сознании, так и в наших информационных системах именно информацию.

С другой стороны, смысл, который тот или иной элемент данных в базе данных имеет для конкретного человека, команды или организации, зависит от соответствия между структурой данных и внутренней когнитивной схемой человека, команды или организации. Если мы утверждаем, что в сознании человека хранится информация (а не данные), то подразумеваем, что признак, отличающий информацию от данных, это прежде всего смысл, а не структура. С этой точки зрения все, что хранится в информационных системах, – это данные.

В литературе можно встретить различные взгляды на роли структуры и смысла при преобразовании данных в информацию. Ясной общепринятой позиции по этому вопросу пока не выработано.

В ряде работ определены основные процессы, с помощью которых данные превращаются в информацию:

- классификация;

⁶⁹ Обсуждение особенностей явных и неявных знаний будет продолжено в главе 3 в связи с рассмотрением концепции динамических способностей и взгляда на фирму, основанного на знаниях.

- перестановка/сортировка;
- агрегирование;
- выполнение вычислений;
- отбор.

При этом важно еще раз отметить, что именно получатель данных определяет, является сообщение данными или информацией. Смысл данных часто определяется через ассоциации с опытом или установление связей с другими данными. Понятие смысла субъективно, и то, что один человек считает ценной информацией, другой может воспринимать как данные, не имеющие особого значения.

2. Преобразование информации в знания.

Итак, информация обычно рассматривается как данные, обработанные таким образом, чтобы быть значимыми (имеющими смысл), ценными и подходящими для конкретной цели. В то же время знания рассматриваются как информация, применимая на практике (имеющая практическое значение), или информация в сочетании с пониманием и возможностями. Но, поскольку в основе определений информации лежит смысл, а для его выявления необходимо понимание, использовать практическую применимость или понимание в качестве признаков, отличающих знания от информации, кажется затруднительным.

Кроме того, различие явного знания и информации кажется еще менее оправданным. Если знания являются достоянием людей и воплощают в себе предшествующее понимание, опыт и результаты обучения, то трудно утверждать, что явные знания, зафиксированные в документах и информационных системах, чем-то отличаются от информации.

Часто выделяют следующие процессы, преобразующие информацию в знания:

- объединение информации из многих источников в течение определенного времени;
- структурирование убеждений;
- изучение предмета и приобретение опыта;
- организация и обработка информации для передачи понимания, опыта, накопленных результатов обучения и компетенции;
- интернализация⁷⁰ с обращением к внутренним когнитивным структурам.

3. Преобразование знаний в мудрость.

Если попытаться сформулировать суммарное определение мудрости на основе приведенных выше соображений, то можно сказать, что это способность действовать наиболее подходящим образом с учетом того, что известно (знания) и что приносит наибольшую пользу (социально-этические нормы).

Мы уже говорили, что мудрость – еще более неуловимое и расплывчатое понятие, чем знания. Оно сильно связано с человеческой интуицией, пониманием, интерпретацией и действиями. Поэтому в рамках данной книги, посвященной прежде всего управлению данными, углубляться в эти связи мы не будем. Отметим только, что важность вклада знаний (а следовательно, данных и информации) в формирование мудрости сомнений не вызывает.

На основе проведенного обсуждения можно сделать следующие выводы⁷¹.

1. Иерархия DIKW в явном виде упоминается не всегда, но она подразумевается в определениях данных, информации, знаний и мудрости в большинстве источников. Обычно информация определяется в терминах данных, знания – в терминах информации, а мудрость – в терминах знаний.

2. В описании процессов, преобразующих элементы, находящиеся ниже в иерархии, в элементы, стоящие над ними, наблюдается меньшая согласованность, и из этого вытекает недостаточная четкость определений. В частности, в отношении следующих аспектов:

⁷⁰ Интернализация (от лат. *interims* – «внутренний») – термин применяется в социологии, психологии, педагогике и культурологии для обозначения процесса освоения индивидом или группой людей социальных ценностей, норм, установок, стереотипов, принадлежащих тем, с кем он, она или они взаимодействуют.

В результате интернализации структуры, внешние по отношению к данной личности или группе, превращаются в их внутренние регуляторы поведения. Механизмы интернализации очень сложны и пока слабо изучены.

⁷¹ Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 2007, 33(2), 163–180. DOI: 10.1177/0165551506070706. – URL: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>.

– Поскольку в качестве признаков, отличающих данные от информации, выделяются структура и смысл, точки зрения на то, хранится информация в информационных системах и в умах людей, или она образуется только в умах, могут расходиться⁷².

– Определения информации (как данных, обработанных для того, чтобы быть значимыми, ценными и подходящими для конкретной цели) и знаний (как информации, имеющей практическое значение) частично совпадают. В связи с этим вопрос соотношения этих понятий нуждается в дальнейшем исследовании.

– Явные знания по своей сути практически не отличаются от информации.

Можно выделить следующие основные характеристики элементов иерархии DIKW, значения которых изменяются при переходе с уровня на уровень:

- ценность;
- смысл;
- структура;
- широта применения;
- возможность передачи;
- необходимость участия человека при вводе (возможность автоматического ввода);
- возможность программируемой обработки;
- возможность передачи.

Изменение значений характеристик отражено на рисунке 2.3.

Уровни пирамиды знаний могут быть сопоставлены с классами информационных систем, которые преимущественно используются для работы с ними (рис. 2.4).

⁷² Это замечание, в частности, показывает, что по сравнению с термином «управление информацией» схожий с ним термин «управление данными» более общий и в некоторых случаях его использование более корректно.



Рис. 2.3.

Изменение значений характеристик элементов иерархии DIKW*



Рис. 2.4.

Сопоставление уровней пирамиды знаний с классами информационных систем*

* Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 2007, 33(2), 163–180. DOI: 10.1177/0165551506070706. – URL: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>.

2.4. Разрыв между данными и информацией

Хотя пирамида знаний служит общепринятой устоявшейся моделью, с каждым годом появляются новые публикации с предложениями по ее корректировке на основе учета современных тенденций.

Можно обратить внимание, например, на статью⁷³. В ней справедливо отмечается, что, хотя общее количество собираемых данных стремительно возрастает, не все из них образуют базу для информации, не говоря уже о знаниях или мудрости. Нельзя не учитывать тот факт, что данные могут быть неточными или ложными. Таким образом, постоянно растущие собрания больших массивов данных также обязательно включают в себя и постоянно растущие собрания больших массивов неточных или ложных данных. На сегодняшний день нет никаких известных исследований, показывающих, остается ли удельный вес неточных или ложных данных постоянным по мере роста общего объема собираемых данных. Возможно, его величина сохраняется прежней или уменьшается, но не исключено, что доля неточных и ложных данных увеличивается. Это может происходить, в частности, из-за стремительного роста различного рода фейковых новостей или же по причине отрицательной реакции людей на сам сбор данных и т. п.⁷⁴ Поэтому просто сбор данных не приносит особой пользы. Что действительно ценно и необходимо, так это увеличение сбора точных и достоверных данных. Приведенные соображения, естественно, предполагают, что конечная цель получения данных – обретение знаний и мудрости, и не распространяются, например, на компании, чья основная деятельность – построение и обслуживание дата-центров, собирающих и хранящих любые данные.

Исходя из этого, в статье предложена нелинейная схема, отражающая взаимоотношения элементов иерархии DIKW (рис. 2.5). На ней элементы представлены в виде перекрывающихся и соприкасающихся областей (диаграмма Венна). Данные и информация не лежат в основе знаний и мудрости, а просто частично входят в их состав. В ближайшие годы область данных, вероятно, будет расти экспоненциально, но еще неизвестно, увеличатся ли в размерах какие-либо другие области. Преимущество приведенной диаграммы в том, что она точнее отражает соотношение представленных на ней понятий, чем пирамида знаний, и, что не менее важно, ориентирует на углубление знаний и обретение мудрости, а не просто на увеличение сбора данных.

⁷³ Van Meter, Heather J. Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom. *Law Technology and Humans*, 2020, Vol. 2. No. 2, 69–80. DOI: 10.5204/lthj.1470. – URL: <https://lthj.qut.edu.au/article/view/1470>.

⁷⁴ Очевидно, с этим связано появление упомянутых нами во введении публикаций о токсичности данных и о том, что данные – это «новый мусор».



Рис. 2.5.

Модернизированная диаграмма DIKW*

* Van Meter, Heather J. Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom. Law Technology and Humans, 2020, Vol. 2. No. 2, 69–80. DOI: 10.5204/lthj.1470. – URL: <https://lthj.qut.edu.au/article/view/1470>.

Основываясь на сведениях, полученных в ходе нашего обзора взаимоотношений между элементами иерархии DIKW, можно сказать, что с точки зрения этих взаимоотношений деятельность любой организации представляет собой множество циклических цепочек преобразований «данные – информация – знания»⁷⁵⁷⁶⁷⁷.

Данные появляются в результате выполнения каких-либо действий, например операций по продаже клиенту товара или предоставлению гражданину государственной услуги (рис. 2.6). Эти данные могут быть преобразованы в ценную информацию, в частности, если зафиксировать сведения о том, что для определенного клиента заказ выполнен повторно. В свою очередь эта информация наполняет хранилище, содержащее корпоративные знания и позволяющее на основании совокупности сведений о заказах сделать, допустим, следующий вывод: клиенты старше 40 лет лучше реагируют на недавно проведенную рекламную кампанию.

⁷⁵ Мы не включаем в эту цепочку «мудрость», поскольку, как уже отмечалось выше в этой главе, «мудрость» – еще более неуловимое и расплывчатое понятие, чем «знания». Оно сильно связано с человеческой интуицией, пониманием, интерпретацией и действиями. Поэтому в рамках данной книги, посвященной прежде всего управлению данными, углубляться в эти связи мы не будем.

⁷⁶ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

⁷⁷ Redman T. C. Data Driven: Profiting from Your Most Important Business Asset. Harvard Business Review Press, 2008.



Рис. 2.6.

Цикл преобразований «данные — информация — знания»

С учетом этих знаний организация может приступить к дальнейшим действиям, например провести целевую кампанию, направленную на охват клиентов в возрасте до 40 лет, что приведет к увеличению продаж. Соответственно, появятся новые данные, новая информация и новые знания, на основе которых будут осуществляться новые действия. Таким образом, цикл повторяется.

В этом цикле данные играют роль первичного строительного блока. Они возникают из действий и приводят к новым действиям. Эффективность преобразования данных в информацию определяет эффективность получения знаний, а также предпринимаемых на их основе дальнейших действий.

В этой связи область бизнес-аналитики (Business Intelligence, BI), обсуждаемую более подробно в следующих главах, можно представить как «завод по очистке данных» (рис. 2.7). Он превращает сырье (данные) в разнообразные информационные продукты: «информацию», которая собирается и агрегируется в хранилищах данных; «знания», которые собираются из запросов, сообщений и от аналитических инструментов; «планы», которые собираются по кусочкам из правил, закономерностей, моделей и схем, обнаруженных с помощью аналитических инструментов; и «действия», посредством которых бизнес-пользователи реализуют планы, генерирующие события, которые в свою очередь начинают новый цикл⁷⁸.

⁷⁸ Эккерсон У. Панели индикаторов как инструмент управления: ключевые показатели эффективности, мониторинг деятельности, оценка результатов / Пер. с англ. — М.: Альпина Бизнес Букс, 2007.

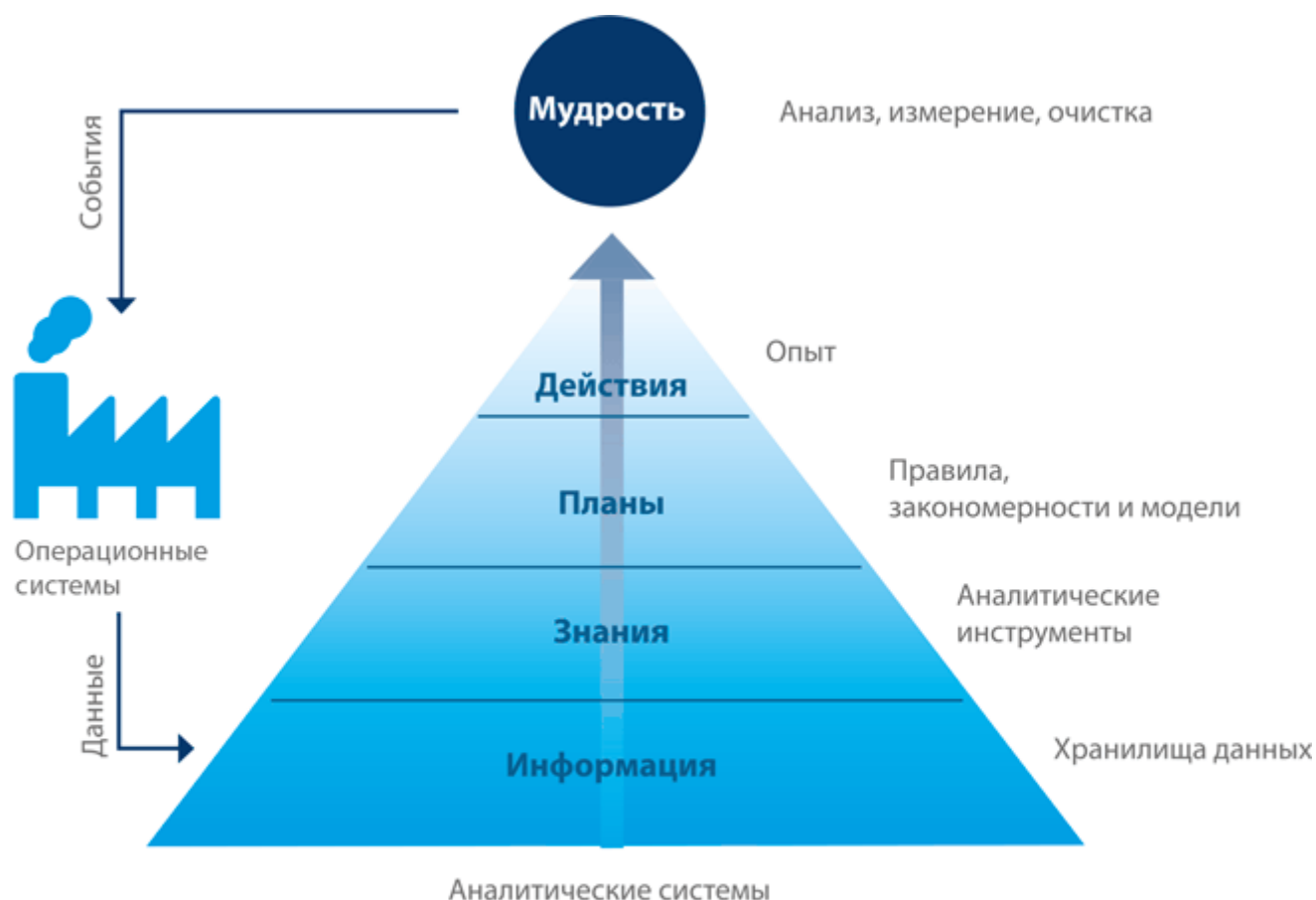


Рис. 2.7.

«Завод по очистке данных»*

* Эккерсон У. Панели индикаторов как инструмент управления: ключевые показатели эффективности, мониторинг деятельности, оценка результатов / Пер. с англ. – М.: Альпина Бизнес Букс, 2007.

Между тем практика автоматизации большинства организаций показывает, что преобразование данных в информацию на сегодня слабое звено в описанной циклической цепочке. Очень часто мы наблюдаем своего рода разрыв (gap) между данными и информацией (см. рис. 2.6). Причиной могут быть данные следующих категорий: плохо определенные, разрозненные, дублирующиеся, низкого качества или устаревшие. Кроме того, данных может быть недостаточно, или же, наоборот, слишком много. Устранение указанных слабых мест, повышение ценности данных – актуальнейшая задача для организации в современных условиях, если она хочет выжить и процветать⁷⁹.

Необходимым шагом в преодолении разрыва между данными и информацией является изменение отношения организации к своим данным. Она должна перестать относиться к данным как к побочному продукту (byproduct) операционной деятельности и начать рассматривать данные как стратегически важный ресурс^{80,81}. Об этом мы поговорим в следующей главе.

В завершение этой главы обратим внимание на следующее. Обсуждая иерархию DIKW и ликвидацию

⁷⁹ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

⁸⁰ Там же.

⁸¹ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Д. Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

разрыва между данными и информацией, мы рассматриваем эти понятия как разные вещи, в действительности они тесно переплетены и по отдельности друг от друга не существуют. Данные – это форма информации, а информация – это форма данных, что, в частности, отмечается в DAMA-DMBOK⁸².

При этом в DAMA-DMBOK подчеркивается, что внутри организации полезно проводить четкую границу между информацией и данными хотя бы в целях более ясного донесения требований и ожиданий по различным направлениям практической работы до различных заинтересованных аудиторий. (Пример: «Предлагаем ознакомиться с отчетом о продажах за минувший квартал [информация]. Он составлен на основе данных нашего информационного хранилища [данные]. В следующем квартале эти результаты [данные] будут использованы для создания сравнительных показателей нашей работы по отношению к предыдущему кварталу [информация]».)

Признание различий между данными и информацией, а также между целями, для которых они могут быть использованы, служит основой стержневого постулата об управлении данными: предметом управления являются и данные, и информация; при этом качество и того и другого возрастает лишь при согласованном управлении с учетом потребностей конечных потребителей. Таким образом, когда мы говорим о преодолении разрыва между данными и информацией, мы, по сути дела, подразумеваем плавный («бесшовный») переход одного в другое. Поэтому в дальнейшем в этой книге, как и в DAMA-DMBOK, термины «информация» и «данные» используются как взаимозаменяемые синонимы (за исключением тех случаев, когда это оговаривается особо).

Глава 3. От побочного продукта к стратегическому ресурсу

3.1. Эволюция концепций стратегического менеджмента

В предыдущей главе мы обратили внимание на то, что необходимый шаг в преодолении разрыва между данными и информацией – изменение отношения к данным. Организация должна перестать относиться к своим данным как к побочному продукту операционной деятельности и начать рассматривать их как один из наиболее важных ресурсов. Далее мы увидим, что в основе такого представления должна лежать стратегия организации – ключевой элемент управления бизнесом. Более того, мы будем подробно рассматривать тесно связанные с ней стратегию работы с данными и стратегию управления данными⁸³. Для более глубокого понимания обсуждаемых далее вопросов целесообразно немного остановиться на понятиях «стратегия организации» и «стратегический менеджмент».

Как пишет Роберт Грант⁸⁴ в своем знаменитом учебнике «Современный стратегический анализ»⁸⁵, стратегия (на самом общем уровне) – это планирование способов, с помощью которых организация или индивид могут достичь поставленных целей. Как только мы переходим от общих понятий стратегии к более точному определению, то все начинает зависеть от того, на какой именно арене реализуется стратегия. Если говорить о войне, то стратегия заключается в том, чтобы одержать военную победу над врагом; если о политике, то речь пойдет о власти положения и поддержке электората, которая позволяет приобрести эту власть и удержать ее в своих руках; если о бизнесе, то стратегия обеспечивает выживание и процветание фирмы.

Учебный курс по стратегическому управлению бизнесом (политике бизнеса, или стратегии) давно

⁸² DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁸³ См. главу «Руководство данными» (глава 10).

⁸⁴ Роберт Грант (Robert M. Grant) – профессор менеджмента Джорджтаунского университета (США), один из ведущих мировых специалистов в области стратегического управления.

⁸⁵ Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

стал обязательным для аккредитованных программ МВА, а зарождение этой дисциплины связано с процессом становления и роста в эпоху промышленного подъема конца XIX – начала XX века (наиболее массовым образом в США) крупных промышленных корпораций или, в терминологии А. Чандлера⁸⁶, «современного делового предприятия».

Задачи профессионализации управления новым типом организаций стимулировали создание в ведущих университетах США школ бизнеса, а вскоре и создание нового учебного курса, призванного интегрировать знания, получаемые студентами в функциональных курсах аккаунтинга, управления продажами и управления производством. В 1911 году в Гарвардской школе бизнеса впервые в мире началось преподавание курса, названного «Политика бизнеса» и ставшего прообразом нынешней учебной дисциплины стратегического управления⁸⁷.

В 1960-е годы вышло несколько ключевых работ, в которых был предложен ряд ставших классическими определений и концепций в области стратегического управления: «Стратегия и структура» А. Чандлера⁸⁸, коллективный учебник Гарвардской школы бизнеса «Политика бизнеса: Текст и кейсы»⁸⁹ и работа И. Ансоффа⁹⁰ «Корпоративная стратегия»⁹¹.

В книге «Стратегия и структура» впервые были серьезно изучены процессы формирования крупных компаний. Описывая истории роста и административных изменений в General Motors, Sears, Standard Oil of New Jersey (Exxon) и DuPont, Чандлер показал, как их руководители выполняют работу по стратегическому управлению и достигают выдающихся результатов. Он утверждал, что решение о стратегии предшествует выбору адекватной ей организационной структуры. Ранее никто не рассматривал стратегию под таким углом. В начале 1960-х годов этот тезис быстро получил признание. Чандлеру принадлежит авторство термина «организационная инновация», описывающего случай разработки и внедрения нетривиальной организационной структуры (например, дивизиональной), востребованной для поддержки новой стратегии (диверсификации).

В «Стратегии и структуре» введен и концептуализирован термин «стратегия» вместо привычного термина «политика бизнеса». Эта книга оказала влияние и на авторов двух других названных классических трудов⁹².

Чандлер определил стратегию как «установление базовых долгосрочных целей и задач предприятия и выработку программы действий и распределения ресурсов, необходимой для реализации этих целей». Это определение активно используется до сих пор, хотя есть множество других, по-разному расставляющих акценты в отношении отдельных составляющих стратегии в зависимости от предпочтений авторов.

Например,⁹³:

86 Альфред Чандлер (Alfred D. Chandler) – профессор истории бизнеса в Гарвардской школе бизнеса; автор ряда известных книг о масштабе действий и структуре современных корпораций.

87 Каткало В. С. Теория стратегического управления: этапы развития и основные парадигмы // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2002, 2: 3–21; 3: 3–26. – URL: <https://cyberleninka.ru/article/n/teoriya-strategicheskogo-upravleniya-etapy-razvitiya-i-osnovnye-paradigmy-2>, <https://cyberleninka.ru/article/n/teoriya-strategicheskogo-upravleniya-etapy-razvitiya-i-osnovnye-paradigmy-3>.

88 Chandler A. D., Jr. *Strategy and structure*: Chapters in the history of American enterprise. Cambridge, 1962.

89 *Business policy: Texts and cases* / Ed. by E. A. Learned, C. R. Christensen, K. R. Andrews, W. D. Guth. Homewood, 1965.

90 Игорь Ансофф (Igor Ansoff) – американский математик и экономист российского происхождения, считается родоначальником стратегического менеджмента.

91 Ansoff H. I. *Corporate strategy: An analytical approach to business policy for growth and expansion*. New York, 1965.

92 Там же.

93 Каталевский Д. Ю. Эволюция концепций стратегического менеджмента: от Гарвардской школы внешней среды до ресурсного подхода к управлению // Государственное управление. Электронный вестник. Вып. 16. Сентябрь, 2008. – URL:

- Стратегия – это система (или план), интегрирующая основные цели организации, политику и последовательность действий в единое целое. Хорошо сформулированная стратегия помогает распределять и размещать ресурсы организации уникальным и эффективным образом, исходя из сферы компетенции и слабостей компании, предполагаемых изменений окружающей среды и внезапных действий умных противников.

- Стратегия – это принцип организации целей и задач, а также политика и планы для достижения этих целей и решения задач, сформулированные таким способом, чтобы определить, в какой сфере бизнеса находится или должна находиться компания и какой она является или должна являться.

- Чем занимается бизнес-стратегия, можно выразить двумя словами – конкурентным преимуществом. Единственная цель стратегического планирования состоит в том, чтобы помочь компании максимально эффективно добиться значимого превосходства над конкурентами. Корпоративная стратегия, таким образом, подразумевает попытку изменить силы компании в лучшую сторону по сравнению с ее конкурентами самым эффективным способом.

- Стратегия – это модель размещения ресурсов, которая позволяет организациям продолжать или совершенствовать свою деятельность.

Грант определил четыре общих фактора успешных стратегий (рис. 3.1).

- Цели, которые являются простыми, согласованными и долгосрочными.
- Глубокое понимание конкурентной среды.
- Объективная оценка ресурсов с целью эффективного использования сильных сторон и защиты уязвимых мест.
- Эффективное выполнение.

Разработку стратегии фирмы (а в более общем смысле любой другой организации) можно рассматривать как формирование связей между этой фирмой и внешней средой (рис. 3.2).

Фирма обладает тремя совокупностями основных характеристик:

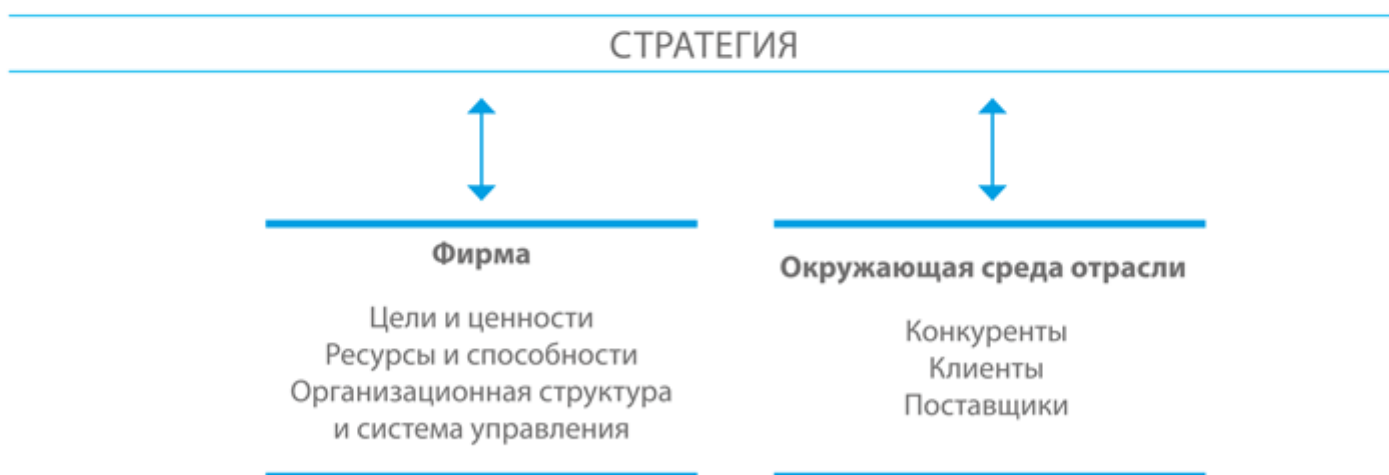
- цели и ценности;
- ресурсы и способности;
- организационная структура и системы управления⁹⁴.

<https://cyberleninka.ru/article/n/evolyutsiya-kontseptsiy-strategicheskogo-menedzhmenta-ot-garvardskoy-shkoly-vneshney-sredy-do-resursnogo-podhoda-k-upravleniyu>.

⁹⁴ Под системами управления здесь понимаются функционирующие в рамках фирмы системы управления отдельными видами ресурсов и функциональными областями (система управления финансами, закупками и т. п.).

**Рис. 3.1.**

Общие факторы успешных стратегий*

**Рис. 3.2.**

Стратегия как связь между организацией и окружающей средой*

* Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

Внешняя среда фирмы включает в себя широкий диапазон экономических, социальных, политических и технологических факторов, которые влияют на принятие решений и ее деятельность. Однако для большинства стратегических решений ядром внешней среды является отрасль, определяемая отношениями с клиентами, конкурентами и поставщиками.

Задача стратегии состоит в том, чтобы установить, каким образом фирма воспользуется своими ресурсами в пределах окружающей среды (и достигнет своих долгосрочных целей) и как она организует себя, чтобы внедрить эту стратегию.

Таблица 3.1.

Эволюция стратегического менеджмента*

Период	1950-е гг.	1960-е — ранние 1970-е гг.	Поздние 1970-е — середина 1980-х гг.
Преобладающая тема	Бюджетное планирование и контроль	Корпоративное планирование	Позиционирование
Основные проблемы	Финансовый контроль	Планирование роста, особенно диверсификация и портфельное планирование	Выбор отраслей и рынков. Позиционирование лидера на рынке
Основные концепции и методы	Составление финансового бюджета. Планирование инвестиций. Оценка проектов	Прогнозирование на длительные и средние периоды. Методы корпоративного планирования. Синергия	Анализ отрасли. Сегментирование. Кривые опыта. Анализ влияния прибыли на рыночную стратегию. Стратегические бизнес-единицы. Портфельное планирование
Организационные следствия	Системы составления операционного бюджета и капитальных смет становятся ключевыми механизмами координации и контроля	Создание отделов корпоративного планирования и долгосрочное планирование процессов. Слияния и поглощения	Мультидивизиональные и мультинациональные структуры. Большая отраслевая и рыночная селективность

* Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

Чтобы представить стратегию как связь между фирмой и ее внешней средой, необходимо воспользоваться фундаментальным понятием стратегического соответствия. Стратегия станет успешной, если она будет соответствовать особенностям внешней и внутренней среды фирмы – целям и ценностям, ресурсам и способностям, организационной структуре и системам управления⁹⁵.

Грант выделяет несколько этапов эволюции стратегического менеджмента, особенности которых

95 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

определялись практическими потребностями бизнеса (табл. 3.1). В каждый из периодов акцент делался на тот или иной фактор из числа выделенных ранее (рис. 3.1) факторов успешных стратегий. Схожая картина эволюции отражена, например, в работах⁹⁶ и⁹⁷.

Приведем описание этапов эволюции стратегического менеджмента из учебника «Современный стратегический анализ».

«...В 1950–1960-е годы руководители компаний стали испытывать большие трудности при согласовании решений и сохранении контроля над компаниями, которые быстро росли и постоянно усложнялись. Составление финансовых бюджетов заложило фундамент для ежегодного финансового планирования, а составление смет капитальных вложений на основе дисконтированного потока наличности дало новый подход к оценке индивидуальных инвестиционных проектов. Корпоративное планирование было разработано в виде системы, позволяющей координировать индивидуальные решения относительно инвестиций капитала и планировать долгосрочное развитие фирмы. Основой нового корпоративного планирования стали макроэкономические прогнозы основных экономических агрегированных показателей, которые затем декомпозировали на прогнозы для отдельных рынков и конкретных товаров фирмы. Типичным форматом стал пятилетний корпоративный план, который устанавливал цели и задачи, приоритеты различных товаров, продуктов и сфер деятельности фирмы, прогнозировал ключевые экономические тенденции (включая рыночный спрос, рыночную долю компании, доход, затраты и прибыли) и размещал капиталовложения.

...В 1960-е и начале 1970-х годов корпоративное планирование делало основной акцент на планировании диверсификации – экспансии в новые секторы бизнеса, часто посредством поглощения.

...В 1970-х годах обстоятельства изменились. Мало того что стратегия диверсификации оказалась не в состоянии обеспечить предвкушаемую синергию, так еще нефтяные кризисы 1974 и 1979 годов возвестили о новой эре макроэкономической неустойчивости и обострении международной конкуренции со стороны развивающихся японских, европейских и юго-восточных азиатских фирм. Столкнувшись с усилением беспокойности окружающей среды, фирмы более не могли планировать на три – пять лет свои инвестиции, вывод на рынок новых продуктов и услуг и количество требуемого персонала просто потому что они не могли предсказать свое столь далекое будущее.

В результате в целях максимизации потенциала прибыльности произошло смещение акцента с планирования на создание стратегии, где детальному менеджменту путей развития компании стали уделять меньше внимания, чем позиционированию компании на рынке относительно конкурентов. Переход от корпоративного планирования к тому, что стало называться стратегическим менеджментом, был связан с усилением внимания к конкуренции как центральной характеристике окружающей среды и конкурентным преимуществам компании как главной стратегической цели.

Этот сдвиг интереса в сторону стратегии, обусловленный поисками способов повысить эффективность работы, сфокусировал внимание на источниках прибыльности. В конце 1970-х и в 1980-х годах акцент переместился на источники прибыли во внешней среде.

...К 1990-м годам основной акцент в анализе стратегии сместился с источников прибыли во внешней среде на источники прибыли внутри самой фирмы. Все чаще и чаще ресурсы и способности фирмы стали рассматривать как основной источник конкурентного преимущества и фундамент для формулирования стратегии. Особое значение, которое придавалось тому, что получило название “ресурсный подход к фирме” (resource-based view of the firm), знаменовало существенный сдвиг в изучении стратегии. Анализ отраслевой структуры подталкивает фирмы к поиску привлекательных рынков и благоприятного стратегического позиционирования. Результатом стало широкое имитирование компаниями стратегий друг друга. Отныне первенство отдавали внутренним ресурсам и способностям: фирмы стали все чаще

96 Каткало В. С. Эволюция теории стратегического управления: монография. – СПб.: Издательский дом Санкт-Петербургского государственного университета, 2006.

97 Каталевский Д. Ю. Эволюция концепций стратегического менеджмента: от Гарвардской школы внешней среды до ресурсного подхода к управлению // Государственное управление. Электронный вестник. Вып. 16. Сентябрь, 2008. – URL: <https://cyberleninka.ru/article/n/evolyutsiya-kontseptsii-strategicheskogo-menedzhmenta-ot-garvardskoy-shkoly-vneshney-sredy-do-resursnogo-podhoda-k-upravleniyu>.

обращать внимание на то, что отличает их от конкурентов, и начали создавать стратегии, использующие эти различия, чтобы обеспечить себе уникальное положение за счет своих конкурентных преимуществ.

Технологический бум конца 1990-х годов стал причиной того, что многие преуспевающие фирмы вновь задумались о своих бизнес-стратегиях, и это при условии того, что большая часть компаний не пережила технологического спада 2000–2002 годов. Быстро сокращающиеся затраты на коммуникации и обработку информации оживили интерес к сетевой экономике и движущим силам войн стандартов, воздействию технологий прорыва, ведущей роли знания и феномену, в соответствии с которым «победитель забирает все рынки». Быстрый темп изменений на рынках технологий стимулировал интерес к применению теории выбора и синергии при разработке стратегии. Наибольший интерес стали вызывать стратегические инновации, поиск новых подходов к получению дохода и созданию конкурентного преимущества в условиях стремительных и непредсказуемых изменений в отраслях промышленности.

...Главным аспектом этих поисков стал интерес к новым моделям бизнеса как фундаментально новым методам получения доступа к источникам стоимости»⁹⁸.

Яркое проявление тенденции устойчивого следования компаний курсом стратегических инноваций – активно проводимая большинством из них цифровая трансформация. Чтобы понять роль данных в этих процессах и составить четкое представление о важности повышения ценности данных для их интенсификации, необходимо более подробно ознакомиться с концепцией ресурсного подхода к фирме.

3.2. Ресурсный подход к фирме

Как было отмечено в предыдущем разделе, фирмы при выработке стратегии лишь совсем недавно стали уделять внимание своим ресурсам и способностям, делая до этого основной акцент на возможности получения прибыли, которые возникают во внешней среде.

В 1990-е годы идеи относительно роли ресурсов и способностей как фундамента стратегии фирмы и главного источника прибыли вылились в единый подход, названный впоследствии «ресурсный подход к фирме» (resource-based view of the firm)⁹⁹¹⁰⁰¹⁰¹¹⁰²¹⁰³¹⁰⁴. Основная идея ресурсного подхода заключается в том, что фирма, по сути дела, представляет собой совокупность ресурсов и способностей, которые являются первичными определяющими факторами ее стратегии и эффективной деятельности¹⁰⁵.

98 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

99 Barney J. B. Firm Resources and Sustained Competitive Advantage // Journal of Management, 1991, 17: 99–120. DOI: 10.1016/S0742-3322(00)17018-4. – URL: <https://www.scribd.com/document/50535995/Barney-J-B-1991-Firm-Resources-and-Sustained-Competitive-Advantage>.

100 Barney J. B. Looking inside for competitive advantage // Academy of Management Executive, 1995, 9 (4): 49–61. DOI: 10.5465/AME.1995.9512032192. – URL: https://d1b10bmlvqabco.cloudfront.net/attach/j7stfd10zcx6hr/j81pk3yhtwx36f/j8p7nfta16wz/Looking_Inside_for_Competitive_Advantage.pdf.

101 Barney J. B. Resource-based theories of competitive advantage: a ten year retrospective on the resource-based view // Journal of Management, 2001, 27: 643–650. DOI: 10.1177/014920630102700602. – URL: <https://www.homeworkmarket.com/sites/default/files/article.pdf>.

102 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

103 Каткало В. С. Ресурсная концепция стратегического управления: генезис основных идей и понятий // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2002, 4: 20–42. – URL: <https://cyberleninka.ru/article/n/resursnaya-kontseptsiya-strategicheskogo-upravleniya-genezis-osnovnyh-idey-i-ponyatiy-1>.

104 Каткало В. С. Место и роль ресурсной концепции в развитии теории стратегического управления // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2003, 3: 3–17. – URL: <https://cyberleninka.ru/article/n/mesto-i-rol-resursnoy-kontseptsii-v-razvitii-teorii-strategicheskogo-upravleniya-predislovie-k-razdelu>.

105 Считается, что первое исследование в этой области принадлежит профессору Высшей школы бизнеса Мичиганского

В основе ресурсного подхода лежат две предпосылки. Согласно первой, организации внутри одной отрасли значительно отличаются друг от друга в отношении обладания доступными ими ресурсами и контроля за ними.

Вторая предпосылка заключается в том, что ресурсы не мобильны, и поэтому ресурсная гетерогенность различных фирм внутри одной отрасли может сохраняться достаточно долгое время. Некоторые ресурсы не могут с легкостью передаваться от одной фирмы к другой, а также между фирмами и рынком. Подобные ресурсы, равно как и компетенции, созданные при их использовании, были названы «липкими» (sticky). Создание долгосрочного конкурентного преимущества зависит от тех специфических особенностей фирмы и от тех ресурсов, которые не так легко передать¹⁰⁶.

Ресурсный подход становится особенно полезным, когда внешняя среда непостоянна, что, как уже отмечалось, характерно для современной ситуации. Тогда сама фирма (обладающая некой совокупностью ресурсов и способностей), а не ее внешняя среда обеспечивает намного более устойчивое основание для определения своей же собственной идентичности. Таким образом, определение фирмы с точки зрения того, что она способна делать, обеспечивает более основательный фундамент для стратегии, чем определение, базирующееся на потребностях, которые фирма стремится удовлетворить.

Ресурсный подход оказал глубокое воздействие на то, как компании формируют свои стратегии. Когда первоочередной задачей стратегии являются выбор отрасли и позиционирование в ней, компании стремятся следовать похожим стратегиям. Напротив, ресурсный подход подчеркивает уникальность каждой компании и утверждает, что секрет прибыльности не в том, чтобы делать то же самое, что все остальные компании, а в том, чтобы использовать собственные отличия. Создание конкурентного преимущества означает формирование и реализацию стратегии, которая опиралась бы на уникальность портфеля ресурсов и способностей каждой фирмы.

Основное положение этого подхода – признание того факта, что фирма должна стремиться к всестороннему и глубокому пониманию своих ресурсов и способностей. Это закладывает фундамент для:

- выработки стратегии, которая позволит эффективно использовать основные сильные стороны организации;
- развития ресурсов и способностей фирмы.

Нужно четко понимать различие между ресурсами фирмы и ее способностями (capabilities¹⁰⁷). Ресурсы – это производственные средства, принадлежащие фирме, а способности – то, что фирма может делать. По отдельности ресурсы не создают конкурентного преимущества; они должны быть задействованы все вместе, только тогда они формируют так называемую организационную способность – способность фирмы осуществлять специфическую производственную деятельность. Именно она является

университета Биргеру Вернерфельту (Birger Wernerfelt), который в статье «Ресурсная трактовка фирмы»^{*} указал на «полезность анализа фирмы, скорее с точки зрения ее ресурсов, чем продуктов», и высказал гипотезу о том, что такой подход может стать новой парадигмой теории стратегического управления^{**}.

^{*} Wernerfelt B. A. A resource-based view of the firm // Strategic Management J. 1984, Vol. 5 (2). P. 17 1.

^{**} Катъкало В. С. Ресурсная концепция стратегического управления: генезис основных идей и понятий // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2 002, 4: 20–42. – URL: <https://cyberleninka.ru/article/n/resursnaya-kontseptsiastrategicheskogo-upravleniya-genezis-osnovnyh-idey-i-ponyatiy-1>.

106 Каталевский Д. Ю. Эволюция концепций стратегического менеджмента: от Гарвардской школы внешней среды до ресурсного подхода к управлению // Государственное управление. Электронный вестник. Вып. 16. Сентябрь, 2008. – URL: <https://cyberleninka.ru/article/n/evolyutsiya-kontseptsiy-strategicheskogo-menedzhmenta-ot-garvardskoy-shkoly-vneshney-sredy-do-resursnogo-podhoda-k-upravleniyu>.

107 Термин capabilities переводится в русских изданиях как «способности» и «возможности». Однако, как отмечает один из ведущих российских экспертов в области стратегического управления В. С. Катъкало, первый вариант представляется единственно правильным. Контраргумент, что понятие «способности» ассоциируется с живыми существами, не выдерживает критики: в словарях русского языка одно из приемлемых толкований его смыслового содержания охотно иллюстрируется как «пропускная способность железных дорог». Кроме того, для знатоков теории стратегического управления термин «возможности» имеет английский эквивалент opportunities из модели SWOT-анализа^{*}.

^{*} Катъкало В. С. Место и роль ресурсной концепции в развитии теории стратегического управления // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2003, 3: 3–17. – URL: <https://cyberleninka.ru/article/n/mesto-i-rol-resursnoy-kontseptsii-v-razvitiy-teorii-strategicheskogo-upravleniya-predislovie-k-razdelu>.

сутью эффективной деятельности. Как пишет Р. Грант, «нейрохирург бесполезен, если рядом нет рентгенолога, анестезиолога, медсестер, хирургических инструментов, видеоаппаратуры и массы других ресурсов. Чтобы поставленная задача была решена, все ресурсы должны работать согласованно»¹⁰⁸.

Организационная способность требует, чтобы опыт разных людей был объединен с основным оборудованием, технологией и другими ресурсами. Такая интеграция осуществляется посредством шаблонных организационных процедур (выполняемых регулярно и предсказуемым образом в соответствии с заведенным порядком как повседневная рутина) – организационных рутин (organizational routines)¹⁰⁹.

Наиболее часто выделяют три основных типа ресурсов: материальные, нематериальные и человеческие. Корпоративный бухгалтерский баланс предоставляет ограниченный перечень ресурсов фирмы, который включает в основном материальные ресурсы. В то же время в большинстве компаний нематериальные ресурсы гораздо сильнее влияют на общую стоимость активов, чем материальные. Таким образом, на основании баланса нельзя составить полное представление о ресурсном потенциале фирмы. Поэтому в последнее время часто обсуждается необходимость пересмотра бухгалтерских балансов, поскольку они, как правило, скрывают стратегически важную информацию, недооценивая или переоценивая активы¹¹⁰.

В связи с темой нашей книги важно подчеркнуть, что предложения по поводу пересмотра бухгалтерских балансов все чаще поступают в отношении такого вида нематериальных ресурсов, как информационные ресурсы¹¹¹¹¹².

Взаимосвязь между ресурсами, способностями и конкурентным преимуществом отражена на рисунке 3.3¹¹³.

Чтобы получить конкурентное преимущество на основе ресурса или способности, необходимы два условия:

- уникальность – ресурс или способность не должны быть широкодоступны внутри отрасли;
- уместность – ресурс или способность должны иметь отношение к одному или нескольким ключевым факторам успеха на рынке (они должны помогать фирме создавать ценности для покупателей или выживать в конкурентной борьбе).

108 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

109 Понятие «рутина» мы более подробно рассмотрим далее в этой главе при обсуждении знаний организации и, в частности, групповых знаний.

110 Там же.

111 Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

112 Мы поговорим об этом более подробно в главе 5.

113 Различные исследователи, занимающиеся проблематикой ресурсной концепции, немного по-разному трактуют такие ключевые понятия, как «ресурсы», «способности» и т. д. (хотя явных расхождений в понимании нет). В этой книге мы придерживаемся трактовки Гранта, который, по мнению многих специалистов, представил категориальный аппарат данной концепции наиболее четко и последовательно*.

* Катяло В. С. Ресурсная концепция стратегического управления: генезис основных идей и понятий // Вестник Санкт-Петербургского университета. Серия «Менеджмент», 2002, 4: 20–42. – URL: <https://cyberleninka.ru/article/n/resursnaya-kontseptsiya-strategicheskogo-upravleniya-genezis-osnovnyh-idey-i-ponyatiy-1>.

Сысоева Е. А. Конкурентные преимущества предприятия: эволюция и источники формирования // Вопросы экономики и права, 2011, 35: 47–51. – URL: <http://ecsocman.hse.ru/data/2012/04/16/1271937531/8.pdf>.



Рис. 3.3.

Взаимосвязь между ресурсами, способностями и конкурентным преимуществом*

* Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

При этом важно первостепенное внимание уделять тем способностям, которые являются основными для стратегии и эффективной деятельности фирмы. Их называют ключевыми компетенциями. Отличие ключевых компетенций от остальных организационных способностей в том, что они:

- вносят непропорционально большой вклад в окончательную потребительскую ценность или эффективность создания этой ценности;
- обеспечивают основу для проникновения на новые рынки¹¹⁴.

3.3. Концепция динамических способностей

Хотя ресурсный подход и является в настоящее время доминирующим в мировой теории и практике стратегического управления, он не избежал критики. В частности, некоторые исследователи считали, что организации вообще нет необходимости разрабатывать конкурентное преимущество, поскольку внешняя среда настолько турбулентна и так быстро меняется, что любое преимущество будет быстро уменьшаться. В связи с этим в рамках ресурсного подхода стала развиваться концепция динамических способностей (dynamic capabilities), которая постепенно становится самостоятельной областью исследований¹¹⁵¹¹⁶¹¹⁷.

¹¹⁴ Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

¹¹⁵ Каткало В. С. Эволюция теории стратегического управления: монография. – СПб.: Издательский дом Санкт-Петербургского государственного университета, 2006.

¹¹⁶ Каталевский Д. Ю. Эволюция концепций стратегического менеджмента: от Гарвардской школы внешней среды до ресурсного подхода к управлению // Государственное управление. Электронный вестник. Вып. 16. Сентябрь, 2008. – URL: <https://cyberleninka.ru/article/n/evolyutsiya-kontseptsii-strategicheskogo-menedzhmenta-ot-garvardskoy-shkoly-vneshney-sredy-do-resursnogo-podhoda-k-upravleniyu>.

¹¹⁷ Шарников А. В. Концепция динамических способностей как основа устойчивости конкурентных преимуществ организации в условиях нестабильности внешней среды // Бизнес. Образование. Право. Вестник Волгоградского института бизнеса, 2012, 4 (21): 184–189. – URL: <http://vestnik.volbi.ru/webarchive/421/yekonomicheskie-nauki/koncepcija-dinamicheskikh>.

Концепция динамических способностей фирмы была сформулирована в 1997 году Дэвидом Тисом¹¹⁸ с соавторами в статье «Динамические способности и стратегический менеджмент»¹¹⁹.

Тис и коллеги определили динамические способности как специфические «организационные процедуры по использованию ресурсов, чтобы соответствовать запросам рынка и даже проводить рыночные изменения». С помощью таких процедур менеджеры влияют на ресурсную базу своей организации – приобретают ресурсы либо избавляются от них, интегрируют и заново комбинируют – для создания новых стратегий. Динамические способности представляют собой основные инструменты по созданию, изменению и рекомбинации совокупных ресурсов в новые источники конкурентного преимущества. Это сложившаяся организационная практика, благодаря которой компании достигают новых конфигураций ресурсов для соответствия изменениям на рынке.

К динамическим способностям фирмы можно отнести такие организационные процедуры и практики, как создание новых продуктов, заключение альянсов и партнерств между компаниями, методики принятия стратегически важных для организации решений и т. п. – все, что помогает создавать дополнительную ценность компании путем манипулирования доступными ей ресурсами.

Некоторые динамические способности нацелены на интегрирование имеющихся у организации ресурсов, как, например, разработка новых продуктов в компании, когда менеджеры объединяют отдельные навыки и знания для разработки новых прибыльных продуктов и услуг. Точно так же к динамическим способностям относится принятие стратегических решений, когда менеджеры на основе своей профессиональной, личной и отраслевой экспертизы принимают решения, которые могут значительно повлиять на судьбу всей компании в целом.

Другие динамические способности направлены на получение новых ресурсов, в частности организационные процедуры и практики в области создания новых знаний, особенно важные в фармацевтических, ИТ и других наукоемких отраслях. Не менее важны организационные способности по заключению альянсов и приобретению новых активов, приносящих организации новые ресурсы и развивающие таким образом ее ресурсную базу. Наконец, к динамическим способностям относится и способность организаций избавляться от ресурсов или их комбинаций, которые стали бесполезными для создания конкурентного преимущества в связи с изменившимися условиями рынка.

В рамках концепции динамических способностей удалось связать ресурсный подход с влиянием внешней окружающей среды, преодолев таким образом некоторую односторонность ресурсного подхода, сосредоточенного на микроуровне организации. Динамические способности зависят в первую очередь от условий внешней среды, рынка, на котором работают компании. Именно рынок определяет организационные процедуры: там, где внешняя среда изменяется не так быстро и структура отрасли стабильна, организационные способности представляют собой традиционно распространенные процессы. Это преимущественно сложные аналитические операции, очень детальные, опирающиеся в основном на существующие знания и накопленный опыт и пользующиеся линейными моделями экстраполяции прошлых результатов для получения прогнозов о будущей динамике. На турбулентных и стремительно меняющихся рынках с нечеткой структурой компании, наоборот, используют простые экспериментальные практики и процедуры, быстро разрабатываемые при получении нового знания и взаимодействии с различными рыночными контрагентами. Такого рода стратегии высокоадаптивны, но сложно предсказуемы по последствиям.

Хотя динамические способности существенно расширили ресурсный подход, они не опровергают его базовые положения о том, что в основе конкурентного преимущества организаций лежат ресурсы.

sposobnostei-ka.html.

¹¹⁸ Дэвид Тис (David J. Teece) – профессор Школы бизнеса им. У. А. Хааса Калифорнийского университета в Беркли. В 2020 году был признан самым цитируемым в мире ученым в области стратегического менеджмента. Статья «Динамические способности и стратегический менеджмент» была в период с 1997 по 2005 год самой цитируемой в мире статьей по экономике и бизнесу.

¹¹⁹ Teece D., Pisano G., Shuen A. Dynamic capabilities and strategic management // Strategic Management Journal, 1997, 18 (7): 509–533. – <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0266%28199708%2918%3A7%3C509%3A%3AAID-SMJ882%3E3.0.CO%3B2-Z>.

Динамические способности – обязательное (но не исчерпывающее) условие получения конкурентного преимущества. Их умелое использование позволяет усилить имеющиеся у организаций комбинации ресурсов¹²⁰.

В 2008 году американские ученые Эль Сави (El Sawy) и Павлоу (Pavlou) в статье «Поддерживаемые ИТ бизнес-способности для турбулентных сред»¹²¹ разработали схему, отражающую соотношение основных видов организационных способностей компании. В наибольшей степени такое соотношение проявляется для компаний, функционирующих в турбулентной среде – быстро меняющейся хозяйственной среде с высоким уровнем неопределенности и непредсказуемости. В статье выделена «триада бизнес-способностей» (рис. 3.4).

- Операционные способности – планомерно реализуемые способности, позволяющие эффективно выполнять повседневные операции в рамках основных видов деятельности, таких как производство, логистика, продажи и т. п.

- Динамические способности – планомерно реализуемые способности, позволяющие эффективно реконфигурировать существующие операционные способности с целью обеспечения соответствия изменениям в бизнес-среде.

- Импровизационные способности – приобретаемые в ходе деятельности компании (и реализуемые непланомерно) способности, позволяющие спонтанно реконфигурировать существующие ресурсы в режиме реального времени, чтобы сформировать новые операционные способности, которые лучше соответствуют новым ситуациям в бизнес-среде.

Исследование Эль Сави и Павлоу показало, что в умеренно турбулентных средах основными способностями, обеспечивающими реконфигурацию существующих операционных способностей, являются динамические, в то время как в очень бурных средах на первый план выходят импровизационные способности (хотя точный переломный момент в исследовании не определен). При этом менеджеры часто отказываются от импровизации (даже при наличии успешной практики), поскольку с точки зрения общепринятой корпоративной культуры импровизация свидетельствует о недостатке планирования. Таким образом, можно сказать, что в современных условиях интенсивных изменений хозяйственной среды динамическим способностям в триаде бизнес-способностей принадлежит главенствующая роль.

Для динамических способностей ученые предложили модель, выделяющую четыре измерения:

- ощущение и восприятие хозяйственной среды;
- освоение современных навыков;
- интегрирование знаний;
- координация (упорядочение и контроль) деятельности.

Модель динамических способностей отражена на рисунке 3.5.

Применительно к теме нашей книги в приведенной модели важно выделить измерение «интегрирование знаний» – встраивание новых знаний в новые операционные способности.

3.4. Взгляд на фирму, основанный на знаниях

В течение последних десятилетий представления о ресурсах, способностях и управлении ими расширились и приняли новое направление благодаря интересу к управлению знаниями. Управление знаниями – это процессы и методы, с помощью которых организации генерируют ценность на основе знания. Уровень интереса к управлению знаниями отмечается в большинстве крупных корпораций. Даже появилась новая должность – директор по управлению знаниями; кроме того, возникло огромное количество консалтинговых фирм, занимающихся управлением знаниями.

¹²⁰ Каталевский Д. Ю. Эволюция концепций стратегического менеджмента: от Гарвардской школы внешней среды до ресурсного подхода к управлению // Государственное управление. Электронный вестник. Вып. 16. Сентябрь, 2008. – URL: <https://cyberleninka.ru/article/n/evolyutsiya-kontseptsiiy-strategicheskogo-menedzhmenta-ot-garvardskoy-shkoly-vneshney-sredy-do-resursnogo-podhoda-k-upravleniyu>.

¹²¹ Sawy O. A., Pavlou P. A. IT-enabled business capabilities for turbulent environments. MIS Quarterly Executive, 2008, vol. 7, no. 3, pp. 57–68. – URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2742325.

В академических кругах интерес к роли знаний в организациях проявляется в слиянии нескольких направлений исследований, включая теорию ресурсов, экономику информации, эпистемологию, эволюционную экономику и управление технологиями. В результате возник «взгляд на фирму, основанный на знаниях», который рассматривает организацию как совокупность знаний, направленных на создание ценности¹²²¹²³¹²⁴¹²⁵.

122 Концепция взгляда на фирму, основанного на знаниях (как расширение концепции ресурсного подхода к фирме), была предложена Робертом Грантом*, чтобы обозначить важную роль знаний как ресурса, который может обеспечить высокую отдачу в течение длительного времени**.

* Grant R. M. Toward a Knowledge-Based Theory of the Firm // Strategic Management Journal, 1996, 17, Winter Special Issue. Pp. 109–122.

** Evans N., Price J. Enterprise information asset management: the roles and responsibilities of executive boards // Knowledge Management Research & Practice, 2016, 14: 353–361. – URL: <https://www.experiencematters.com.au/wp-content/uploads/2018/03/Enterpriseinformation-asset-management-the-roles-and-responsibilities-of-executive-boards.pdf>.

123 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

124 Grant R. M. Toward a Knowledge-Based Theory of the Firm // Strategic Management Journal, 1996, 17, Winter Special Issue. Pp. 109–122.

125 Nickerson J., Zenger T. A knowledge-based theory of the firm: the problem-solving perspective // Organization Science, 2004, 15 (6), 617–632.



Рис. 3.4.

Триада бизнес-способностей*



Рис. 3.5.

Модель динамических способностей**

* Sawy O. A., Pavlou P. A. IT-enabled business capabilities for turbulent environments. MIS Quarterly Executive, 2008, vol. 7, no. 3, pp. 57–68. – URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2742325.

** Никифорова Ю. М. Динамические способности как необходимый инструмент для нового развития и управления компаний // Современная конкуренция, 2019, т. 13, 10 (59): 73–89. – URL:

<https://cyberleninka.ru/article/n/dinamicheskie-sposobnosti-kakneobhodimyy-instrument-dlya-novogo-razvitiya-i-upravleniya-kompaniy>.

Управление знаниями и взгляд на фирму, основанный на знаниях, – важное продолжение анализа ресурсов и способностей. С точки зрения ресурсов знания – самый важный производительный ресурс. Ценность людей и механизмов заключается в том, что они воплощают в себе знания. Со стратегической точки зрения знания – это наиболее интересный ресурс: многие виды знаний представляют собой большую редкость, огромная часть знаний с трудом поддается передаче, а сложные формы знаний очень трудно копировать. Способности можно рассматривать как проявление знаний организации. Инструменты управления знаниями способны сделать более ясной структуру организационной способности и указать, каким образом можно эту способность формировать, развивать, сохранять и воспроизводить¹²⁶.

Как уже отмечалось в главе 2, с точки зрения эпистемологии (теории познания) знания могут быть неявными (неформализованным) и явными (формализованным).

Неявные знания не формализуются и могут существовать лишь вместе с их обладателями – отдельными людьми или группами лиц. Неявные знания сложно передаются от одного человека к другому. Они связаны с интуицией, прозрением, догадками, идеалами, ценностями. Этот вид знаний – основа индивидуальных действий и опыта.

Явные знания выражены в виде слов и цифр и могут передаваться в формализованном виде на физических носителях.

Организации не только и не столько сортируют, хранят и используют имеющиеся знания, сколько создают новые знания в процессе инновационной деятельности, в том числе опираясь на неявное знание. Каждый работник изначально включен в процесс создания нового знания организации. Японские исследователи И. Нонака (Ikujiro Nonaka) и Х. Такеучи (Hirotaka Takeuchi) предложили концепцию компании, создающей знание. Эта концепция предусматривает использование в равной мере явного и неявного знания в процессе постоянного наращивания интеллектуального капитала¹²⁷.

Знания формирует опыт. Информация, полученная из учебных курсов, бесед, книг, наслаиваясь на реальный опыт, может формировать новое знание. Опыт открывает историческую перспективу для оценки новых ситуаций и событий, поэтому знания иногда коррелируются с опытом. Когда организация нанимает эксперта, она фактически покупает его оценки, базирующиеся на его личном опыте.

Носителем и явного, и неявного знания может быть не только конкретная личность, но и организация. Следовательно, можно говорить и о неявном групповом знании, которое лежит в основе устойчивых моделей коллективных реакций и внутренних взаимодействий.

В литературе для обозначения неявного группового знания иногда используется термин «рутины», который обозначает повторяющиеся типовые действия, регулярные поведенческие шаблоны организации. Рутин – то, что происходит автоматически, без инструкций и в отсутствие процедуры выбора, при этом рутины не могут быть кодифицированы¹²⁸.

В русском языке под рутиной понимается заведенный порядок, установившаяся практика, определенный режим, шаблон, сложившиеся правила, касающиеся занятий людей.

Таким образом, персональное неявное знание – это прежде всего умения. В то же время групповое неявное знание – это рутины. Рутин существуют не изолированно, образуя взаимную зависимость. Некоторые рутины могут быть неявными для одних членов организации и явными для других¹²⁹.

В главе 2 мы уже говорили о том, что знания существуют в континууме между неявными знаниями (знать как) и явными знаниями (знать что). Явные знания – это неявные знания, которые были

126 Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

127 Нонака И., Такеучи Х. Компания – создатель знания. Зарождение и развитие инноваций в японских фирмах. – М.: Олимп-Бизнес, 2011.

128 Понятие «организационная рутина» уже рассматривалось в этой главе при обсуждении организационных способностей.

129 Гапоненко А. Л., Орлова Т. М. Управление знаниями: как превратить знания в капитал. – М.: Эксмо, 2008.

задокументированы. Отношение к явным и неявным знаниям со стороны коммерческих организаций весьма противоречиво. Множество организаций стремятся перевести неявные знания в явные, чтобы не зависеть от отдельных личностей и продублировать значимые достижения. В то же время эти организации не заинтересованы в том, чтобы основные конкурентные преимущества перешли в форму, готовую для дублирования. Именно поэтому они стараются сохранять некоторые из своих конкурентных преимуществ в тех формах, которые не поддаются дублированию (специфические тренинги, корпоративная культура, специальные системы обслуживания и пр.). Таким образом, границы между явными и неявными знаниями относительно, более того, можно утверждать о степени неявности тех или иных знаний¹³⁰.

Первоначально управление знаниями занималось главным образом информационной технологией, в частности использованием интранета (внутренних сетей), средствами автоматизации групповых процессов и базами данных для хранения, анализа и распространения информации (т. е. управлением явными знаниями¹³¹)¹³². Последующее развитие управления знаниями привело к тому, что данным и информации стали уделять все меньше внимания, а организационному обучению (т. е. управлению неявными знаниями) – все больше, особенно передаче лучших практик и управлению интеллектуальной собственностью¹³³.

Управление знаниями в целом не является предметом рассмотрения этой книги. Далее мы будем обсуждать только его первую составляющую – управление данными. Как уже отмечалось ранее в этом разделе, с точки зрения ресурсов знания – самый важный производительный ресурс организации. Данные (явные знания) являются важной составной частью знаний и, следовательно, так же самым важным (наряду с неявными знаниями) производительным ресурсом.

В завершение данного раздела следует отметить известную статью Майкла Портера¹³⁴ (с соавтором) «Каким образом информация дает вам конкурентное преимущество»¹³⁵. В ней подчеркивается важность данных и информации как производительного ресурса и описываются основные механизмы их влияния на цепочку ценности и формирования конкурентных преимуществ на их основе¹³⁶.

В следующем разделе мы обсудим ресурсоориентированный взгляд на данные и информацию, обеспечивающий всесторонний учет основных особенностей этого ресурса с целью извлечения из него максимальной выгоды.

Использование данных для обеспечения ключевых компетенций и динамических способностей (примеры бизнес-потребностей и решений)

Машиностроение

• Анализ бизнес-потребностей:

– Сокращение количества случаев упущенной выгоды в результате незаключения контрактов на поставку продукции из-за отсутствия точной, полной и актуальной информации об

¹³⁰ Там же.

¹³¹ В главе 2 уже отмечалось, что явные знания по своей сути практически не отличаются от информации.

¹³² Lambe P. J. The Unacknowledged Parentage of Knowledge Management // Journal of Knowledge Management, 2011, 15 (2): 175–197. – URL: https://www.researchgate.net/publication/220363236_The_Unacknowledged_Parentage_of_Knowledge_Management.

¹³³ Грант Р. М. Современный стратегический анализ. 5-е изд. / Пер. с англ., под ред. В. Н. Фунтова. – СПб.: Питер, 2008.

¹³⁴ Майкл Портер (Michael Porter) – известный американский экономист, специалист в области изучения экономической конкуренции, автор концепции цепочки ценности.

¹³⁵ Porter M., Millar V. How information gives you competitive advantage // Harvard Business Review, 1985, 63 (4): 149–160. – URL: <https://palfreymanventures.net/porter-and-millar-1985.pdf>.

¹³⁶ Понятие «цепочка ценности» (value chain) будет подробно рассмотрено в главе 7.

ассортименте выпускаемых изделий.

– Сокращение количества невыполненных (выполненных с нарушением сроков) заявок на поставку запчастей в процессе сервисного обслуживания изделий из-за неоднозначной идентификации номенклатурных позиций.

● **Решение:** внедрение единого каталога продукции корпорации (объединение локальных каталогов).

● **Анализ бизнес-потребностей:**

– Снижение трудозатрат на проектирование и производство, исключение случаев порчи оснастки и брака продукции из-за несоответствия данных при передаче из конструкторских бюро на заводы электронных макетов изделий.

– Повышение оперативности и снижение погрешностей при подготовке сводных планов и отчетов из-за несоответствия номенклатурных позиций в различных информационных системах.

– Снижение затрат на закупки материально-технических ресурсов и стоимости изделий из-за ошибок в описаниях и дублирования позиций в документах на закупки.

● **Решение:** внедрение единых номенклатурных справочников на уровне холдинговых компаний и организаций корпорации: покупные и комплектующие изделия, стандартные изделия, материалы, услуги (консолидация локальных справочников).

Электроэнергетика

● **Анализ бизнес-потребностей:**

– Снижение затрат из-за рассогласований в описании оборудования в бухучете, производственном учете, АСУТП, диспетчерских системах.

● **Решение:** внедрение единого справочника активов корпорации (консолидация локальных справочников).

Общепромышленные эффекты

● **Анализ бизнес-потребностей:**

– Повышение оперативности и уменьшение погрешностей при подготовке сводной отчетности по организациям, входящим в состав корпорации.

– Повышение достоверности аналитических данных, подготавливаемых в разрезе организаций и контрагентов для выработки управленческих решений.

● **Решение:** внедрение единого справочника оргструктуры и единого справочника контрагентов корпорации и ее организаций (консолидация локальных справочников).

3.5. Данные как ресурс: ресурсоориентированный взгляд на данные и информацию

Одной из первых серьезных публикаций, появившихся в результате формирования взгляда на данные и информацию как на корпоративный ресурс – так называемого ресурсоориентированного взгляда на информацию (resource-based view of information), стал доклад «Управление знаниями как корпоративным ресурсом»¹³⁷, подготовленный в 1976 году для Министерства обороны США¹³⁸. В нем рассматривается подход к управлению знаниями с точки зрения информационных технологий, т. е. прежде всего в контексте управления данными (явными знаниями). Теоретические основы управления корпоративными знаниями в части управления корпоративными данными (в терминах корпоративной архитектуры данных) были заложены еще в 1970-х годах. Формирование взгляда на данные как на корпоративный ресурс было отчасти обусловлено ростом понимания нерациональности создания в компаниях многочисленных независимых хранилищ данных для различных приложений, контент в которых часто дублировался¹³⁹.

¹³⁷ Berry J. F., Cook C. M. Managing Knowledge as a Corporate Resource. Department of Defense, Washington, DC. 1976.

¹³⁸ Об этом говорится, в частности, в обзорной статье «Неподтвержденное происхождение управления знаниями» (The Unacknowledged Parentage of Knowledge Management)*.

* Lambe P. J. The Unacknowledged Parentage of Knowledge Management // Journal of Knowledge Management, 2011, 15 (2): 175–197. – URL: https://www.researchgate.net/publication/220363236_The_Unacknowledged_Parentage_of_Knowledge_Management.

¹³⁹ Lambe P. J. The Unacknowledged Parentage of Knowledge Management // Journal of Knowledge Management, 2011, 15 (2):

Устойчивый рост интереса к ресурсоориентированному взгляду на информацию в США с середины 1970-х годов до начала 1980-х прослеживается в статье «Оценка ценности информации в организациях: вызов для 1980-х»¹⁴⁰. Исследование начинается с момента, когда правительство США осознало угрозу утонуть в работе с бумагами, что повлекло бы неприемлемые затраты. В связи с этим была создана Комиссия по федеральным документам, которая пришла к выводу, что «данными и информацией можно и нужно управлять точно так же, как мы управляем людскими, физическими и финансовыми ресурсами. Данные и информация должны подчиняться тем же принципам в части бюджетной, управленческой и аудиторской практики, что и любые другие ресурсы».

В 1988 году хорошо проработанный подход к реализации управления данными в контексте ресурсоориентированного взгляда был представлен в известной книге Барка (C. Burk) и Хортон (F. Horton) «Информационная карта: Полное руководство по выявлению корпоративных информационных ресурсов»¹⁴¹.

Чтобы правильно управлять любым ресурсом, необходимо понимать, какую роль он играет, его свойства, возможности, которые он предлагает, а также шаги, которые необходимо предпринять, чтобы использовать эти возможности. Но данные не только предоставляют огромные возможности, они также создают огромные проблемы. Например, в отличие от других ресурсов, данные могут быть легко скопированы, переданы многим людям с использованием информационных технологий, а затем использованы десятками различных способов. Но эффективное распространение данных происходит редко. Вместо этого большинство компаний и частных лиц, сознательно или нет, накапливают данные, что часто приводит к жестоким политическим битвам за право собственности. Даже в ситуациях, когда данные являются общими, отдельные лица и подразделения создают и изменяют свои собственные копии. Неизбежные несоответствия в первую очередь подрывают цель обмена данными. Кроме того, организациям трудно обеспечить соблюдение конфиденциальности и других политик, касающихся использования данных.

Какие свойства и особенности данных отличают их от других ресурсов организации? Ответу на этот вопрос посвящена следующая глава.

Литература к главе 3

• Кольке Г. И. Формирование модели конкурентных преимуществ организации с точки зрения ресурсного подхода // Вестник Сибирского института бизнеса и информационных технологий, 2019, 1 (29): 45–52. – URL: <https://cyberleninka.ru/article/n/formirovaniemodeli-konkurentnyh-preimuschestv-organizatsii-s-tochki-zreniya-resursnogo-podhoda>.

• Мичурина О. Ю. Ресурсная концепция фирмы: интегративный аспект // Экономические науки, 2009, 10 (59): 164–168. – URL: https://ecsn.ru/files/pdf/200910/200910_164.pdf.

• Соколов Б. И., Воронов В. С. Институциональные основы информационно-финансового конструирования // Проблемы современной экономики, 2017, 2 (62), 146–151. – URL: <https://cyberleninka.ru/article/n/institutsionalnye-osnovy-informatsionno-finansovogokonstruirovaniy>.

• Evans N., Price J. Barriers to the Effective Deployment of Information Assets: An Executive Management Perspective // Interdisciplinary Journal of Information, 2012, 7: 177–199. – URL: https://www.researchgate.net/publication/289400090_Barriers_to_the_Effective_Deployment_of_Information_Assets_An_Executive_Management_Perspective.

• Levitin A. V., Redman T. C. Data as a resource: Properties, implications, and prescriptions // Sloan Management Review, 1998, 40 (1): 89–101. – URL: <https://sloanreview.mit.edu/article/data-as-a-resource->

¹⁷⁵–197. – URL: https://www.researchgate.net/publication/220363236_The_Unacknowledged_Parentage_of_Knowledge_Management.

¹⁴⁰ Black S. H., Marchand D. Assessing the Value of Information in Organisations: A Challenge for the 1980's // The Information Society Journal, 1982, 1 (3): 191–235.

¹⁴¹ Burk C., Horton F. Infomap: A Complete Guide to Discovering Corporate Information Resources. Englewood Cliffs, NJ: Prentice-Hall.

properties-implications-and-prescriptions/.

- Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark.

- Otto B. Quality and Value of the Data Resource in Large Enterprises // Information Systems Management, 2015, 32 (3): 234–251. DOI: 10.1080/10580530.2015.1044344 – URL: https://www.researchgate.net/publication/275637181_Quality_and_Value_of_the_Data_Resource_in_Large_Enterprises.

- Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira&>;

Глава 4. Данные как ресурс: особенности и подходы к управлению

4.1. Свойства данных как ресурса

Свойства данных и информации как ресурса не раз обсуждались во многих публикациях. Одна из наиболее подробных и обстоятельных – статья американских специалистов Левитина (A. Levitin) и Редмана (T. Redman) «Данные как ресурс: свойства, особенности и рекомендации»¹⁴². Авторы подчеркивают, что в своей работе они рассматривают исключительно данные (а не информацию и знания), потому что считают, что данные заслуживают рассмотрения сами по себе¹⁴³. В то же время многие выработанные ими рекомендации по управлению данными могут дополнить уже известные рекомендации по управлению информацией и знаниями, основанные на лучших практиках. Действительно, компания, которая не может эффективно управлять своими данными, вряд ли справится с работой по управлению своей информацией и знаниями.

Левитин и Редман, ссылаясь на авторитетный американский словарь The American Heritage Dictionary, приводят следующие определения понятия «ресурс» (resource):

- доступный запас, который может быть использован в случае необходимости;
- средство, которое может быть использовано с выгодой.

Таким образом, ресурсы необходимы организации для осуществления ее повседневной деятельности, а также являются потенциальным источником еще нереализованных выгод.

Имея в виду эти определения, мы можем выделить основные категории ресурсов, используемых современными организациями. Существуют традиционные ресурсы – финансовые, человеческие, здания (сооружения) и оборудование, сырье и энергия и так называемые ресурсы информационной эпохи – данные, информация и знания.

Данные можно рассматривать в двух аспектах.

- Данные как таковые – значения атрибутов, характеризующих тот или иной объект определенного типа в соответствии с некоторой моделью данных, описывающей этот тип. Например: тип объекта – сотрудник; модель данных: фамилия, имя, отчество; значения: Иванов, Иван, Иванович.
- Записи данных – физические проявления данных, размещенные в бумажных картотеках, электронных таблицах, базах данных и представляемые пользователям таким образом, чтобы их было легко хранить и использовать.

Остановимся на основных отличительных свойствах данных и записей данных и обсудим, как они соотносятся со свойствами традиционных ресурсов (табл. 4.1).

¹⁴² Levitin A. V., Redman T. C. Data as a resource: Properties, implications, and prescriptions // Sloan Management Review, 1998, 40 (1): 89–101. – URL: <https://sloanreview.mit.edu/article/data-as-a-resource-properties-implications-and-prescriptions/>.

¹⁴³ В следующих главах мы увидим, что, хотя часто понятия «данные» и «информация» не разделяют и считают синонимами, в то же время, когда обсуждается их роль как ресурса, то рассматривается прежде всего аспект данных, а когда обсуждается такое понятие, как «активы», – аспект информации.

Таблица 4.1.

Основные свойства ресурсов организации*

Характеристика \ Ресурс	Финансовые	Человеческие	Здания (сооружения) и оборудование	Сырье	Энергия	Данные	
						Как таковые	Записи
Материальность	См. текст	Да	Да	Да	Да	Нет	Да
Расходуемость	Да	Нет	Нет	Да	Да	Нет	Нет
Возможность совместного использования	Нет	Нет	Нет	Нет	Нет	Да	Да, для записей, хранящихся на компьютерах
Копируемость	Нет	Нет	Нет	Нет	Нет	Неприменимо	Да
Транспортабельность	Быстро	Медленно	Очень медленно / невозможно	Очень медленно	Быстро	Неприменимо	Быстро, для записей, хранящихся на компьютерах
Заменяемость	Да	Да	Да	Да	Да	Нет	См. текст
Недолговечность	Нет	Нет	Нет	Нет	Нет	Неприменимо	Да, для записей, хранящихся на компьютерах

* Levitin A. V., Redman T. C. Data as a resource: Properties, implications, and prescriptions // Sloan Management Review, 1998, 40 (1): 89–101. – URL: <https://sloanreview.mit.edu/article/data-as-a-resource-properties-implications-and-prescriptions/>.

Характеристика \ Ресурс	Финансовые	Человеческие	Здания (сооружения) и оборудование	Сырье	Энергия	Данные	
						Как таковые	Записи
Универсальность	Высокая	Средняя	Низкая	Низкая	Высокая	Средняя	
Оцениваемость	Да	Да	Да	Да	Да	См. текст	
Обесцениваемость						Да, за некоторыми исключениями	
Источник	Внешний/внутренний	Внешний	Внешний	Внешний	Внешний	Внешний/внутренний	
Возобновляемость	Неприменимо	Неприменимо	Неприменимо	Неприменимо	Неприменимо, за исключением солнечной	Да	Да, с задержкой
Компьютерное (электронное) хранилище	Нет	Нет	Нет	Нет	Нет	Нет	Да, для записей, хранящихся на компьютерах

1. Нематериальность

В литературе существует всеобщее согласие, что наиболее яркой характеристикой, отличающей данные от традиционных ресурсов, служит их нематериальность (intangibility). В прошлом нематериальность данных заставляла некоторых исследователей возражать против рассмотрения данных (а также информации и знаний) в качестве ресурсов. Эта точка зрения становилась непопулярной по мере возрастания роли нематериальных ресурсов во всех сферах, от традиционного производства до финансовых услуг, страхования и других отраслей, интенсивно потребляющих информацию.

При этом важно проводить различие между данными как таковыми и представлением данных на определенном носителе. Если рассматривать данные как таковые, то они явно нематериальны. Но записи данных вполне материальны, будь они на бумаге, микрофильмах или компьютерных носителях (например, оптический диск или флеш-накопитель). В некоторых случаях нам требуется специальное оборудование, такое как компьютер, чтобы убедиться в наличии записей данных, но эта потребность не делает их нематериальными.

Финансовые ресурсы (реализация которых представляет собой либо наличные деньги, либо записи данных) в принципе также можно рассматривать как нематериальные. Однако чаще их считают материальными, поскольку во многих случаях (в частности, в финансовой сфере) под материальностью понимают способность оцениваться в денежном выражении.

2. Расходуемость

Важная характеристика любого ресурса – расходуемость (consumability). Ресурс является расходуемым, если использование уменьшает его объем, доступный для использования в будущем. Деньги, сырье и энергия – примеры расходуемых ресурсов. Очевидно, что ни данные как таковые, ни записи данных не являются расходуемыми.

Иногда утверждается, что нерасходуемость данных отличает их от всех остальных более традиционных ресурсов. Но человеческие ресурсы, равно как здания (сооружения) и оборудование, также не подлежат расходованию в соответствии с данным выше определением. Действительно, назначение работника или машины для выполнения конкретной задачи не исключает переназначения после завершения задачи. Что касается вопросов, связанных с износом, то мы их обсудим позже, когда будем рассматривать характеристику «обесцениваемость» (depreciability).

3. Возможность совместного использования

Под возможностью совместного использования (shareability) ресурсов мы подразумеваем возможность того, что несколько пользователей могут одновременно использовать одну и ту же единицу ресурса. Ни один из традиционных ресурсов не является общим в соответствии с этим определением. Например, два пользователя не могут совместно использовать одну и ту же денежную купюру. Они могут делить здание, но не одну и ту же его часть. Однако данные для совместного использования доступны, причем двумя способами. Во-первых, одни и те же данные могут иметь несколько представлений в разных наборах записей, каждый из которых может использоваться одновременно разными пользователями. Это делает данные как таковые общедоступным ресурсом. Во-вторых, современные системы управления базами данных обеспечивают одновременный многопользовательский доступ к одним и тем же записям данных. Однако эта возможность может быть затруднена ограничениями носителя (например, бумажные записи недоступны для одновременного использования). Возможность совместного использования также может быть намеренно ограничена по соображениям безопасности с помощью шифрования или путем ограничения доступа к записям данных с помощью паролей. Совместное использование данных подразумевает как преимущества, так и опасности, которые не имеют отношения к управлению традиционными ресурсами.

4. Копируемость

Еще одно свойство данных – их копируемость (copyability): можно создать идентичную единицу рассматриваемого ресурса за долю стоимости оригинала. В данном определении требование к снижению стоимости имеет решающее значение, поскольку без него мы были бы вынуждены заключить, что оборудование, сырье и энергия также являются копируемыми ресурсами. На самом деле, конечно, копируются записи данных, а не сами данные.

Поскольку компьютерные записи являются как нерасходуемыми, так и совместно используемыми, теоретически нет необходимости их копировать. Более того, дополнительные копии не только требуют дополнительных носителей, но и, что важнее, создают проблему поддержания согласованности. С другой стороны, для копирования есть практические причины. Во-первых, оно позволяет пользователям работать с данными в более удобных условиях. Во-вторых, дает им возможность контролировать свою собственную копию данных, что может быть важным по различным политическим соображениям. В-третьих, копирование может повысить долговечность данных – снизить вероятность их непреднамеренного уничтожения (см. ниже). Определение желаемой степени избыточности – важный вопрос при управлении данными как ресурсом.

5. Транспортабельность

Как бы ни была важна способность к копированию, именно транспортабельность (transportability) положила начало информационному веку, обеспечивая почти мгновенный перенос данных на большие расстояния. Возможность локального копирования данных существовала еще до того, как прогресс в области телекоммуникационных технологий сделал возможной передачу данных между удаленными пунктами. (Если быть точным, электронная передача данных на самом деле является не транспортировкой записей данных, а скорее созданием копий в пункте назначения.)

Эффективность современных телекоммуникаций не ограничивается скоростью передачи данных; качество и экономическая эффективность одинаково впечатляют. За исключением, возможно, электроэнергии, никакие другие ресурсы не могут транспортироваться с такой легкостью и эффективностью, как данные, хранящиеся в электронном виде. Другое возможное исключение – деньги, которые также могут передаваться в электронном виде.

6. Незаменяемость

Заменяемость (fungibility) означает, что одна единица рассматриваемого ресурса может быть заменена другой единицей того же ресурса, если последняя доступна. Деньги, сырье и энергия заменяемы.

Человеческие ресурсы, здания (сооружения) и оборудование также заменяемы, хотя их замена может быть дорогостоящей и неудобной. Но единицы данных (т. е. отдельные элементы данных) уникальны: мы не можем заменить дату рождения человека другим элементом данных (например, элементом «имя» или «пол») об этом человеке или о ком-то другом. Хотя иногда мы можем вывести значение одного элемента данных из значения другого (например, определить возраст на основе даты рождения), такие ситуации являются исключительными. Что касается записей данных, мы, очевидно, можем заменить одну запись другой, если они обе представляют один и тот же элемент данных.

Незаменяемость (nonfungibility) данных вызывает особые проблемы управления. Например, в качестве защиты от возможных дефектов в единицах других ресурсов менеджеры могут выбрать сохранение большего объема рассматриваемого ресурса. Но для данных эта стратегия бессмысленна: дефектный элемент данных не может быть заменен другим элементом данных. И конечно, хранение дополнительных копий записей данных не поможет, если копии сделаны с негодного оригинала или устарели.

7. Недолговечность

Под недолговечностью (fragility) ресурса мы подразумеваем легкость, с которой он может быть непреднамеренно уничтожен или потерян при обычном использовании. Традиционные ресурсы обычно не квалифицируются как недолговечные. Ситуация с данными совершенно иная. Хотя бумажные записи могут быть непреднамеренно потеряны или уничтожены, именно удивительная легкость, с которой записи данных, хранящиеся в компьютере, могут быть непреднамеренно перезаписаны или полностью стерты, заставляет нас считать их недолговечными. Кроме того, оцифрованные данные могут быть случайно уничтожены, когда новые информационные системы заменяют старые, или могут быть легко потеряны среди больших объемов других данных. Конечно, хранящиеся на компьютере данные могут быть защищены от перезаписи и регулярно дублироваться в виде резервных копий. Тем не менее многим пользователям не хватает навыков, необходимых для уверенной работы с такими механизмами защиты. Учитывая недолговечность данных, неудивительно, что многие пользователи склонны превышать меры разумной осторожности. Практика показывает, что неоправданное распространение копий данных часто вызвано беспокойством о недолговечности данных.

8. Универсальность

Универсальный (versatile) ресурс – ресурс, который может быть использован для различных целей. Например, конкретное сырье для конкретного производственного процесса может иметь ограниченное альтернативное применение. С другой стороны, деньги имеют самый широкий спектр возможного использования. Данные занимают среднее положение между этими двумя полюсами.

Универсальность данных, наряду с их другими свойствами, предоставляет организациям ценные источники новых возможностей для развития бизнеса и улучшений. Один из примеров – целевой (targeted) маркетинг, основанный на данных (data driven). Отрицательной стороной универсальности данных является возможность злоупотреблений. Например, когда данные, законно собранные для одной цели, используются для другой, незаконной. Данные о возрасте и состоянии здоровья человека, законно собранные в медицинских целях, не должны влиять на возможности продвижения этого человека по службе. Неоднозначности в семантике данных усугубляют эту проблему. Например, продавец может считать продажу завершенной, когда он и клиент устно договорились о сделке. Но юридический отдел не считает продажу завершенной, пока не подписан контракт, производственный отдел – пока продукт не доставлен, а финансовый отдел – пока не получен платеж.

Практическая важность проблем неверно истолкованной семантики данных часто недооценивается. В результате принимаются неоптимальные, даже крайне неправильные решения. Кроме того, ошибочное толкование семантики данных может привести к конфликтам. Например, все различные интерпретации продажи верны с точки зрения (обычно ограниченной) функциональных подразделений. Поэтому следует ожидать, что они будут энергично защищать свои интерпретации. Это может поставить в тупик компанию, которой необходимо увеличить продажи.

9. Оцениваемость

Под оценкой (valuation) ресурса мы подразумеваем выражение его ценности в денежном эквиваленте. Для традиционных ресурсов ценность определяют либо рыночные силы, либо устоявшаяся практика бухгалтерского учета. Хотя некоторые наборы данных можно купить на открытом рынке (например,

исторические показатели финансовых рынков, данные о точках продаж, многие виды списков клиентов), большинство наборов данных не подлежат продаже, и их оценка создает сложные теоретические и практические проблемы.

Десятилетия исследований этих проблем не дали четких результатов по нескольким причинам: неспособность отделить информационное содержание от информационных технологий, склонность анализировать данные как заменяемый товар или ресурс, отсутствие внутренней (intrinsic) ценности данных (т. е. ценность данных зависит от конкретных приложений), а также универсальность использования данных.

Оценка данных вызывает несколько дополнительных проблем. Во-первых, всегда легче оценить затраты на данные, чем оценить стоимость приносимых ими выгод, что может привести к ошибочным решениям не получать данные, полезность которых вызывает сомнения. Во-вторых, проблемы, связанные с внутренними ценами на передачу данных, могут быть трудноразрешимыми. Например, некоторые организации взимают с пользователей плату за доступ к наборам данных. Такая практика может препятствовать использованию и в итоге противоречить намерениям руководства. С другой стороны, без выяснения готовности пользователей платить трудно отсеять те наборы данных, которые не приносят никакой пользы¹⁴⁴.

10. Обесцениваемость

Обесценивание (depreciation) определяется как уменьшение или потеря стоимости из-за износа, возраста или по другим причинам. Здания, оборудование и большинство сырьевых материалов (кроме энергии) обесцениваются. Потеря стоимости финансовых и людских ресурсов неоднозначна: для первых она осложняется возможностью инфляции и дефляции, для вторых трудности связаны с различными соотношениями между возрастом и производительностью труда, а также с различиями между отдельными работниками.

Ценность данных обычно не уменьшается из-за использования. Но есть и исключения. Чем больше людей используют информацию о состоянии курсов акций на бирже, тем меньше ее ценность для каждого человека. Аналогичным образом, предприятия не продают и не делятся своими данными, поскольку это может снизить их ценность. Во многих случаях имеет значение истечение времени. Если представляют интерес только текущие значения атрибутов (например, размер зарплаты), они должны обновляться по мере изменения аспектов реального мира, которые они описывают. Если данные имеют временные метки, то они, как правило, становятся менее ценными с течением времени (например, данные о зарплате сотрудника за текущий год, вероятно, будут намного менее ценными через десять лет). Но здесь также есть исключения. Методы интеллектуального анализа данных (data-mining) обеспечивают успешное использование подробных исторических записей, поэтому более старые данные могут быть весьма ценными. Таким образом, в большинстве случаев, за некоторыми исключениями, данные не обесцениваются с использованием, но они обесцениваются с течением времени.

11. Множественность источников

Как правило, традиционные ресурсы производятся за пределами организации, использующей их, за исключением финансовых ресурсов, которые могут образовываться как вне, так и внутри компаний. Такая же дихотомия источников происхождения существует и для данных.

В отличие от других ресурсов, данные генерируются огромным количеством источников. Каждая операция с поставщиками и клиентами, большинство внутренних операций, а также управленческая и экспертная деятельность – все это порождает данные. Хотя по отношению к данным применимы такие модели управления, как «потребитель – поставщик», само разнообразие источников данных добавляет массу сложностей.

Часто источники многих наборов данных не документированы или даже неизвестны. Как правило, эти данные не могут быть ни использованы, ни улучшены. Интернет только усугубляет проблему. Конечно, мы не всегда знаем источник других ресурсов, например нескольких литров бензина. Но согласованные стандарты помогают гарантировать, что бензин из разных источников взаимозаменяем. Наличие же у данных такого свойства, как незаменимость, препятствует стандартизации и осложняет работу с

¹⁴⁴ В следующей главе мы подробно рассмотрим основные модели, которые могут быть применены для оценки данных и информации.

источниками¹⁴⁵.

12. Возобновляемость

Всякий раз, когда реальный мир меняется, описывающие его данные также изменяются или создаются новые данные. Новые данные появляются в результате повседневной деятельности организаций или индивидуумов с поразительной скоростью. Это свойство данных, которое можно назвать возобновляемостью (renewability), в гораздо меньшей степени относится к другим ресурсам, за исключением, возможно, солнечной энергии. Другие ресурсы, конечно, могут быть возобновлены. Но спонтанный характер, скорость и степень обновления данных гораздо выше, чем у любого другого ресурса.

Ситуация немного сложнее для записей данных. В большинстве случаев требуется время и усилия, чтобы изменения были отражены. Например, адрес человека меняется в день его переезда, но до обновления базы данных проходит некоторое время. Информационные технологии могут уменьшить время задержки во многих ситуациях, но оно не может быть устранено полностью.

Естественно, менеджеры хотят использовать последние данные. Свойство возобновляемости выдвигает для исполнения этого желания два требования. Во-первых, поскольку обеспечение актуальности данных – важная задача, бизнес-процессы, фиксирующие изменения, должны быть надежными. Во-вторых, механизмы обеспечения актуальности должны быть синхронизированы, иначе в избыточных базах данных обязательно возникнут несоответствия. При неучете этих требований совещания по принятию решений могут перерасти в жаркий обмен мнениями о том, чьи данные точнее.

13. Компьютерное (электронное) хранилище

Данные, в отличие от большинства других ресурсов, могут храниться на компьютерах (в электронных хранилищах). Исключение составляют финансовые ресурсы, которые могут храниться в виде электронных денежных средств. Это свойство вносит свой вклад в обеспечение других свойств, таких как копируемость, возможность совместного использования и транспортабельность. Электронные средства хранения облегчают управление данными, поскольку стоимость хранения данных невелика по сравнению со стоимостью хранения других ресурсов. С другой стороны, дешевое хранилище может способствовать принятию решений о сохранении всего, включая данные, которые больше не нужны. Непреднамеренным последствием является то, что полезные данные становится труднее найти. Дешевое хранилище также может способствовать принятию несколькими подразделениями в рамках организации независимых решений о хранении собственных копий, что приводит к избыточности данных.

В следующем разделе мы обсудим, каким образом специфические свойства данных влияют на подходы к управлению этим ресурсом.

4.2. Особенности управления данными как ресурсом

Управление любым ресурсом направлено на достижение следующих целей:

- иметь достаточный, но не чрезмерный запас ресурса, основанный на потребностях пользователей и моделях использования;
- предоставлять законным пользователям своевременный и эффективный доступ к ресурсу;
- защищать ресурс от незапланированного уничтожения и несанкционированного доступа и использования; поддерживать и улучшать качество ресурса;
- содействовать эффективному использованию ресурса для максимальной выгоды организации.

Для каждой из этих целей в отношении данных можно выделить особенности управления, связанные с рассмотренными в предыдущем разделе фундаментальными свойствами этого ресурса¹⁴⁶.

1. Управление поставками

Основные вопросы управления поставками включают определение пользователей ресурса, их потребностей и моделей использования, а также принятие мер по получению ресурса. Для традиционных ресурсов, таких как сырье, новые технологии позволили применять подход к управлению запасами «точно

¹⁴⁵ Основные виды источников данных будут нами рассмотрены в главе «Данные как объект управления» (глава 12).

¹⁴⁶ Там же.

в срок». Аналогичные подходы могут быть применены и к управлению поставками данных¹⁴⁷ (во избежание создания ненужных данных), но при этом нельзя недооценивать возникающие проблемы.

Прежде всего незаменимость данных подразумевает, что проблема поставок заключается не в том, чтобы верно определить количество необходимых единиц ресурса, а скорее в релевантности данных. Ключевой вопрос для менеджеров не в том, сколько единиц данных необходимо, а в том, каких именно данных.

Во-вторых, в то время как нерасходуемость данных устраняет одну из проблем управления поставками традиционных ресурсов – недостаточность, она создает хорошо знакомую проблему переизбытка данных. Хранение ненужных данных обходится дорого не столько потому, что это приводит к уменьшению свободного места на носителе, сколько потому, что отвлекает внимание руководства и затрудняет поиск необходимых данных. Только лишь получение достоверной информации о различных данных, хранящихся в большой организации, – достаточно сложная задача.

В-третьих, определить пользователей и понять их потребности гораздо сложнее для данных, чем для других ресурсов. Кроме того, часто приходится согласовывать требования к данным разных пользователей (например, различные семантические интерпретации, казалось бы, одинаковых терминов), что также сложная проблема, с которой не сталкиваются традиционные ресурсы.

В-четвертых, некоторые запросы на получение данных являются изменчивыми и непредсказуемыми. Нерасходуемость данных не устраняет потребности знать шаблоны их использования (в частности, с целью обеспечения своевременного обновления значений данных и гарантирования их актуальности).

В-пятых, как упоминалось ранее, стоимость и ценность данных плохо изучены, что затрудняет определение потерь, вызванных отсутствием данных (по сравнению с затратами на поддержание их избыточного объема). Кроме того, для традиционных ресурсов, несмотря на дополнительные расходы и неудобства, как правило, можно удовлетворить неожиданный всплеск спроса (за счет заимствования денег, найма временных работников и т. д.). К данным это не относится.

Наконец, традиционные ресурсы обычно приобретаются под контролем одного подразделения, ответственного за приобретение ресурсов для всей организации. Данные же собираются или производятся отдельными подразделениями для удовлетворения их собственных потребностей практически без централизованного контроля, что облегчается развитием децентрализованных и мобильных вычислений. При этом один из наиболее важных вопросов в управлении поставками ресурсов – выбор поставщиков. Для многих традиционных ресурсов доступны или могут быть легко получены списки существующих альтернатив и критериев выбора (например, цена, качество). Это редко обеспечивается в отношении данных. К счастью, к поставщикам данных могут быть успешно применены общие принципы управления взаимоотношениями между клиентами и поставщиками.

2. Предоставление доступа

Если у организации есть необходимые ресурсы, то пользователи должны иметь к ним доступ. Хотя все понимают трудности с привлечением людей и оборудования для расчистки территории склада после неожиданной метели, справедливо сказать, что вопросы доступа к данным гораздо сложнее, чем для других ресурсов. Проблемы варьируются от архитектуры программных и технических средств для хранения данных и обеспечения доступа к ним до возможности пользователей находить нужные им данные и обмена данными.

В большинстве случаев только данные, хранящиеся в электронном виде, могут удовлетворять требованиям по доступу в современной организации. С учетом прогресса в скорости и доступности электронных накопителей вызывает недоумение, что большие объемы данных все еще хранятся в виде бумажных записей. Возможности по мгновенному переносу записей данных, хранящихся в компьютерах, способствовали концентрации важных для организации сведений в больших базах данных мейнфреймов, доступных пользователям с удаленных терминалов. Какое-то время эта архитектура казалась техническим решением проблемы доступа к данным, однако быстрое распространение персональных компьютеров и новых способов создания сетей привело к созданию клиент-серверных архитектур и распределенных баз данных. Похоже, что как только какая-либо архитектура организации данных получает широкое признание, технический прогресс и новые требования пользователей делают ее устаревшей.

¹⁴⁷ В главе 7 будет подробно рассмотрено понятие «цепочка поставок данных» (data supply chain).

Из этой динамичной трансформации можно вынести несколько уроков. Хотя компьютеры обеспечивают мгновенный доступ к данным, хранящимся в электронном виде, большинство крупных организаций сталкиваются с ошеломляющим разнообразием данных, разработанных для отдельных бизнес-задач, при недостаточной координации решений в части аппаратных средств и программного обеспечения, а также подходов к моделированию данных. Обеспечение эффективного доступа к данным в этих средах хранения и обработки затруднено. Многие пользователи даже не знают, где искать нужные им данные.

Таким образом, первый урок заключается в том, что, пока ресурс данных не будет сформирован (или, что более реалистично, переформирован) в качестве общеорганизационного ресурса, никакое технологическое новшество не гарантирует эффективного доступа к данным организации. Во-вторых, скорость доступа к данным оказалась лишь одним из критериев эффективности. Важны и другие, особенно контроль и гибкость. Это привело к тому, что организации хранят данные, используемые в операционных целях, в транзакционных системах, а данные для поддержки принятия решений – в хранилищах и витринах данных. Следует ожидать, что подобная запланированная сегментация продолжится. И конечно, как число пользователей, так и уровень их требований к данным будут быстро расти. В-третьих, потребителям данных следует ожидать дальнейшего совершенствования технологических средств их хранения и передачи, а также пользовательских интерфейсов. Периодически накапливаемые изменения требований будут приводить к серьезным изменениям в организации данных и доступе к ним.

Совместное использование данных вызывает дополнительные проблемы. Теоретически сочетание свойств данных, включая возможность совместного использования, нерасходуемость и транспортабельность, должно способствовать созданию разнообразных эффективных механизмов совместного использования. Однако на практике их реализация вызывает затруднения из-за множества технических проблем (хотя в целом технические решения имеются). Различного рода политические вопросы еще более запутаны. Как правило, данные собираются и используются подразделениями организации для выполнения своих конкретных задач без учета потребностей других подразделений или организации в целом. Владение данными расширяет влияние. Поэтому мотивация сообщать другим о наличии данных (не говоря уже о том, чтобы делиться ими) не очень высока¹⁴⁸.

3. Безопасность

Забота о безопасности любого ресурса связана с двумя основными проблемами: защитой от незапланированного уничтожения и предотвращением доступа неавторизованных пользователей. Но данные вводят новые проблемы. Во-первых, несанкционированный доступ к важным данным может иметь более серьезные последствия, чем доступ к любому другому ресурсу, даже к деньгам. Во-вторых, хрупкость и невосприимчивость данных делают проблему безопасности особенно острой. В-третьих, нерасходуемость данных снижает вероятность того, что организация обнаружит несанкционированное использование, просто наблюдая, что количество данных уменьшается. Тот факт, что использование данных, хранящихся в электронном виде, не требует их близкого физического расположения, усугубляет ситуацию. В то же время плюсом является то, что копируемость данных предполагает простую стратегию их защиты от возможного уничтожения путем периодического создания резервных копий. Кроме того, существует ряд дополнительных методов защиты данных, включая шифрование данных, программное обеспечение для обнаружения вирусов и программное обеспечение для контроля доступа. Но несмотря на наличие этих инструментов и очевидную важность безопасности, различные исследования показывают недостаточность усилий, направленных на обеспечение безопасности во многих организациях.

Менеджеры должны учитывать вопросы конфиденциальности и приватности. Очевидно, что универсальность данных угрожает конфиденциальности: данные, собранные без каких-либо возражений против возможного нарушения конфиденциальности для одной цели, вполне могут повлечь опасность такого нарушения при других вариантах использования. Существует противоречие между перспективами интеллектуального анализа данных и поиска неизвестных сведений о клиентах и их ожиданиями в отношении приватности. Хотя информация о политике и практике организаций, касающаяся конфиденциальности, доступна, вопросы, связанные с конфиденциальностью и приватностью, нельзя

¹⁴⁸ Более подробно вопросы предоставления доступа к данным обсуждаются в главе «Обеспечение доступности и обслуживание данных: основы» (глава 12).

назвать полностью решенными¹⁴⁹.

4. Качество

Почти для всех организаций качество данных стало одной из основных проблем. Хотя в сфере ИТ акроним GIGO¹⁵⁰ известен уже довольно много лет, особая озабоченность по поводу качества данных как в частном, так и в государственном секторе, возникла лишь недавно. Признание важности качественных данных растет с каждым днем.

Понимание природы данных способствует эффективному управлению качеством данных. Во-первых, обеспечение качества данных включает в себя вопросы, связанные с моделями данных, значениями атрибутов и записями данных. Даже, казалось бы, обыденные решения по моделированию данных могут иметь огромные последствия – достаточно вспомнить трудности и расходы, связанные с «проблемой 2000 года», которая была вызвана решениями сэкономить несколько байтов памяти. Во-вторых, данные абстрактны, поэтому мы не можем определить их качество путем прямого измерения (как, например, можем измерить химический состав сырья). Обычно мы должны сравнивать данные с реальными объектами, которые они описывают, что может быть дорогостоящим или, для некоторых исторических данных, невозможным. В-третьих, из-за незаменимости данных мы не стремимся к единообразию. Поэтому здесь трудно применять стандарты и дефектный элемент данных не может быть просто заменен другим элементом требуемого качества. В-четвертых, из-за таких свойств данных, как возможность совместного использования, копируемость и транспортабельность, дефектные элементы данных могут почти мгновенно достигать большого числа пользователей. В этом отношении некачественные данные подобны вирусу. Невозможно сказать, где они появятся и какое влияние окажут. В-пятых, в отличие от других ресурсов, требуемые уровни качества обычно неизвестны пользователям. Наконец, сам объем обычно создаваемых и хранимых данных усугубляет проблемы управления их качеством.

К счастью, решения большинства из этих проблем находятся и многие организации успешно улучшают качество своих данных. Наиболее эффективные стратегии улучшения, как правило, сосредоточены не на поиске и выявлении ошибок, а на выявлении и устранении их коренных причин. Программы обеспечения качества данных нелегко реализовать, но те организации, где они внедрены, сообщают о хороших результатах: снижении затрат, повышении удовлетворенности клиентов и более уверенном принятии решений¹⁵¹.

5. Управление использованием

Использование любого ресурса наиболее выгодным образом – всегда сложная задача. Что касается традиционных ресурсов, то основная проблема – это их распределение между подразделениями организации на основании запросов на определенные ограниченные объемы. Классический пример – запрос финансов (формирование бюджета). Проблемы оптимизации использования данных совсем иные. Во-первых, многие организации не используют эффективно уже доступные данные. Большинство организаций достаточно эффективно используют данные в операционной деятельности, но гораздо хуже – в планировании и принятии решений (многие организации признают, что они «богаты данными и бедны информацией»¹⁵²).

За последнее десятилетие появилось много новых методов, технологий и инструментов, позволяющих эффективно использовать данные, но только небольшое количество организаций активно применяют их в своих интересах. Даже те, кто добился успеха, использовали лишь малую долю потенциала своих данных.

Вопрос использования становится еще более важным на стратегическом уровне. Фундаментальные

¹⁴⁹ Более подробно вопросы обеспечения безопасности данных обсуждаются в главе «Базовая поддержка жизненного цикла данных» (глава 15).

¹⁵⁰ GIGO (Garbage In, Garbage Out, «Мусор на входе – мусор на выходе»).

¹⁵¹ Более подробно вопросы управления качеством данных обсуждаются в главе «Базовая поддержка жизненного цикла данных» (глава 15).

¹⁵² Эту проблему мы рассматривали в главе 2 (см. раздел «Разрыв между данными и информацией»).

вопросы (какие данные нам понадобятся для выполнения такой стратегии? как мы можем использовать данные для создания новых возможностей? как доступ к новым видам данных повлияет на нашу стратегию?) редко всплывают. При этом вполне вероятно, что рациональное использование данных будет еще более важным условием успеха в будущем по двум ключевым причинам. Во-первых, управление традиционными ресурсами основано на данных, что делает их метаресурсом. Во-вторых, конкуренция в течение предыдущих десятилетий несколько уравнила рыночных игроков с точки зрения использования традиционных ресурсов, выделив данные в качестве потенциально наиболее перспективного источника для получения конкурентных преимуществ¹⁵³.

В конечном счете язык, на котором говорят все менеджеры, – это финансы. Поэтому следует заметить, что в основе большинства проблем использования данных лежит трудность их оценки¹⁵⁴.

6. Дополнительные особенности

Можно выделить еще две дополнительные особенности управления данными.

Во-первых, организационные вопросы вносят свой вклад во многие проблемы, которые мы обсудили выше. Вопросы владения и подотчетности в отношении данных не решены в большинстве организаций. Политическая борьба за контроль над данными и информацией одна из самых упорных. Кроме того, до сих пор нет полного консенсуса по поводу определения соответствующей управленческой инфраструктуры. Хотя современная иерархическая форма управления может не в полной мере подходить для информационной эпохи, она пока сохраняет свои позиции¹⁵⁵.

Во-вторых, ни один другой ресурс не испытывает таких взрывоопасных изменений, как данные, в том, что касается его растущей важности и тех технологических средств, с помощью которых он приобретается, хранится, транспортируется и используется. Это состояние почти непрерывных изменений делает маловероятным полное разрешение всех обозначенных проблем в краткосрочной перспективе.

4.3. Концепция управления информационными ресурсами

Отмеченные выше особенности управления данными как ресурсом привели к постепенному формированию концепции управления информационными ресурсами – Information Resource Management (IRM). IRM рассматривает данные как ресурс организации, который имеет жизненный цикл, подразумевающий создание, распространение, использование и утилизацию. В центре внимания этого подхода – максимизация эффекта, получаемого с помощью данных, по сравнению с затратами на их сопровождение^{156, 157, 158}.

Один из наиболее развитых вариантов реализации IRM был представлен в 1988 году в книге Барка и Хортон «Информационная карта: Полное руководство по выявлению корпоративных информационных ресурсов» (отмеченной в предыдущей главе)¹⁵⁹. Авторы рассмотрели вопросы выявления информационных ресурсов, жизненно важных для деятельности организации. Подход касался информационных ресурсов,

¹⁵³ Тему использования данных для получения конкурентных преимуществ мы рассматривали в предыдущей главе.

¹⁵⁴ Основные модели, которые могут быть применены для оценки данных и информации, обсуждаются в следующей главе.

¹⁵⁵ Организационным вопросам управления данными посвящены главы «Руководство данными» (глава 10) и «Организационные аспекты управления данными» (глава 16).

¹⁵⁶ Black S. H., Marchand D. Assessing the Value of Information in Organisations: A Challenge for the 1980's // The Information Society Journal, 1982, 1 (3): 191–225.

¹⁵⁷ Burk C., Horton F. Infomap: A Complete Guide to Discovering Corporate Information Resources. Englewood Cliffs, NJ: Prentice-Hall, 1988.

¹⁵⁸ Oppenheim C., Stenson J., Wilson R. M. S. The attributes of information as an asset // New Library World, 2001, 10² (11): 458–464. DOI: 10.1108/EUM000000006203 – URL: <https://www.libqual.org/documents/admin/stenson.pdf>.

¹⁵⁹ Burk C., Horton F. Infomap: A Complete Guide to Discovering Corporate Information Resources. Englewood Cliffs, NJ: Prentice-

уже имеющихся в организациях, и определения способов их использования. Информационным ресурсам ставилась в соответствие та или иная ценность на основе стратегических весов (общая бизнес-стратегия организации определяет критерии для взвешивания отдельных информационных ресурсов с точки зрения их полезности для конкретных стратегий). Затраты на поддержание ресурсов рассматривались только после назначения весовых коэффициентов. Было предложено введение в организации регулярной практики оценки полезности информационных ресурсов по отношению к затратам на их поддержание. Регулярный информационный аудит обеспечивал достижение наилучшей отдачи от понесенных затрат. Принципиальное значение имела установленная Барком и Хортоном связь между бизнес-стратегией и информационными ресурсами¹⁶⁰.

Различные авторы по-разному расставляют акценты при описании реализации концепции IRM, тем не менее можно выделить ряд ключевых элементов. Довольно полно такие элементы описаны Левитиным и Редманом в виде рекомендаций для компаний по управлению данными как ресурсом¹⁶¹.

1. Внедрите целенаправленную внутреннюю программу качества данных

Как отмечалось выше, обеспечение качества данных подразумевает учет множества соображений. Два из них наиболее просты: обеспечение четкого определения моделей данных и точности значений данных. Создатели моделей данных и редакторы значений данных (сотрудники, осуществляющие первоначальный ввод и корректировку значений) должны нести ответственность за четкое определение данных и точность. Менеджеры должны снабдить создателей моделей и редакторов данных инструментами управления качеством, включая средства, позволяющие осуществлять анализ потребностей клиентов, измерение, контроль качества, анализ первопричин недостатков. Преимущества получают как создатели моделей и редакторы, так и пользователи данных. Пользователи выигрывают, потому что данные, которые они используют, легче понять и интерпретировать, и они могут использовать их с уверенностью. Создатели моделей и редакторы выигрывают, потому что сокращаются затраты, которые они несут, отвечая на вопросы и внося исправления.

2. Внедрите программу управления поставщиками данных

Такие программы аналогичны внутренним программам обеспечения качества данных, но направлены на улучшение работы с поставщиками данных вне организации. Программы управления поставщиками ресурсов доказали свою огромную эффективность в производстве. Аналогичные программы в отношении данных, должным образом утвержденные, приносят огромную выгоду от поставки нужных данных при низких затратах на внедрение.

Существует множество способов определения и реализации программ управления поставщиками данных. Все они в той или иной степени сосредоточены на качестве данных, поэтому настоящая рекомендация тесно связана с первой.

3. Постоянно уточняйте свои потребности в данных

Большинство организаций имеют гораздо большее количество данных, чем могут использовать, но в то же время у них нет данных, которые им действительно нужны. Зачастую в основе этой проблемы лежит неспособность четко определить, как будут использоваться данные. Гораздо легче просто их собирать. В результате накапливается слишком много данных сомнительной ценности. Для разрешения подобных ситуаций следует подробно определить наиболее важные направления использования данных, перевести эти направления в требования к данным, сообщить об этих требованиях поставщикам данных и исключить из дальнейшего рассмотрения данные, которые не соответствуют требованиям.

4. Определите наиболее важные информационные цепочки и обеспечьте управление ими

Информационные цепочки – это кросс-функциональные бизнес-процессы, в рамках которых

Hall, 1988.

¹⁶⁰ Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

¹⁶¹ Levitin A. V., Redman T. C. Data as a resource: Properties, implications, and prescriptions // Sloan Management Review, 1998, 40 (1): 89–101. – URL: <https://sloanreview.mit.edu/article/data-as-a-resource-properties-implications-and-prescriptions/>.

осуществляется манипулирование необработанными (сырыми) данными (полученными из внешних источников или созданными внутри компании), преобразование их в более ценные данные или информацию, предоставление доступа к ним потребителей и использование их для создания бизнес-ценности. В большинстве организаций управление отдельными функциями и их выполнение осуществляются довольно эффективно. В то же время управление взаимодействием функций реализовано плохо, отчего страдает итоговый результат. Следует определить наиболее важные информационные цепочки и применить к ним методы управления процессами, принятые для обработки данных. Первый шаг – определение кросс-функциональной подотчетности менеджеров.

5. Пересмотрите роль ИТ

ИТ позволяют хорошо налаженным и управляемым информационным цепочкам работать быстрее и дешевле с большей пропускной способностью. Но технология сама по себе или, что еще хуже, технология в сочетании с неэффективным управлением данными не является решением. Чрезмерная зависимость от ИТ в этом случае только усугубляет проблемы. Организации должны сначала наладить управление информационными цепочками и упорядочить их, прежде чем применять новейшие технологии. Нельзя автоматизировать беспорядок.

6. Обеспечьте разработку, ведение и предоставление широкого доступа к реестру наборов данных

Большинство организаций понятия не имеют о том, какие данные у них есть, какие наиболее важные, каковы их источники и степень избыточности. Информационные цепочки – это источники большого количества данных, и они могут быть наиболее важным элементом реестра наборов данных. Разработка полного реестра может оказаться очень сложной задачей и не стоить затраченных усилий. Но организация в состоянии по крайней мере разрабатывать, обновлять и публиковать перечень наиболее важных наборов данных и их источников. Целесообразно начать решать эту задачу с внедрения процесса, который захватывает новые данные и определяет их источники. Как только этот процесс будет реализован, можно перейти к внедрению процесса идентификации и добавления существующих данных.

7. Определите условия и правила, в соответствии с которыми одно подразделение организации может иметь доступ к данным другого подразделения

Большинство организаций и частных лиц неохотно делятся данными. Естественно, существует множество обстоятельств, при которых данные не должны передаваться. Но в общем случае они не должны утаиваться. Четко оговоренные условия и правила должны рассматривать все аспекты разрешенного использования данных: случаи перепродажи данных организацией, предоставляющей доступ; справедливую цену передачи; технические средства доступа; обеспечение качества; средства разрешения конфликтов и т. п.

8. Учитывайте возможность политических противостояний, которые невозможно выиграть, и предотвращайте их

Примером может служить, казалось бы, безобидная задача определения понятия «клиент», которое будет использоваться во всей компании. Такое определение может потребоваться, например, для облегчения ведения бизнеса с этой компанией или для выявления возможностей перекрестной продажи продуктов и услуг. Но у разных подразделений несколько различающиеся отношения с клиентами. Накопленный опыт и знания для ведения бизнеса отражены в модели данных каждого подразделения. И эта модель данных стала формальным определением клиента. Задача разработки общего определения эквивалентна тому, чтобы попросить каждое подразделение снизить эффективность своего способа работы с клиентами. Ни одно подразделение не согласится на это, если только новый способ не будет явно лучше.

9. Четко определите ответственность менеджмента за данные

Возможно, было бы целесообразно включить эти обязанности в общую политику в области данных. Поскольку ни одна организация не может реализовать перечисленные здесь рекомендации одновременно, политика должна развиваться по мере реализации отдельных рекомендаций.

10. Обеспечьте возможности по участию в руководстве программой управления данными менеджеров высшего звена

Описанные проблемы достаточно сложны. Плохие данные могут распространяться как вирус, в то время как хорошие данные остаются запертыми в местах хранения. Обсуждаемые здесь рекомендации требуют, чтобы все менеджеры и подразделения выполняли свою часть работы. Без сильного руководства

со стороны топ-менеджмента программы управления данными рискуют стать жертвой разрозненных планов деятельности функциональных блоков и их руководителей.

Каждая из рассмотренных рекомендаций требует строгой подотчетности менеджмента за данные. В этой связи одна из распространенных ошибок заключается в предположении, что если данные находятся в компьютерах, то ответственность за них должен нести директор по ИТ. В то время как ИТ-директор может нести ответственность за базовую технологию, большая часть данных генерируется и используется операционными и принимающими решения подразделениями, которые не находятся под его контролем. ИТ-директор в соответствии со своей сферой ответственности часто не осведомлен о важных данных организации. Ответственность за них должна лежать на тех, кто находится ближе к созданию и использованию данных.

Выделенные выше рекомендации 5 и 8 (роль технологий и предотвращение политических противостояний) зависят от индивидуальной инициативы. Обеспечение выполнения рекомендаций 1–3 (программа качества, программа управления поставщиками и уточнение потребностей в данных) обычно должно входить в сферу ответственности функциональных блоков, хотя общие инструменты и методы может предоставлять централизованная функция. Рекомендация 4 (информационные цепочки) требует кросс-функционального согласования. Обеспечение выполнения рекомендации 6 (реестр наборов данных) – обычно централизованная функция. Рекомендация 7 (условия и правила обмена данными) требует согласия подразделений, владеющих данными.

Данные создаются и используются в огромных количествах в различных видах деятельности, начиная от самой обыденной операции и заканчивая самым далеко идущим стратегическим решением. Ими редко хорошо управляют. Чтобы улучшить эту ситуацию, организации и менеджеры должны понимать свойства данных и управлять ими как ресурсами. Как мы уже обсуждали в начале этой главы, данные, в отличие от многих обычных ресурсов, нематериальны, легко копируются и транспортируются, а также возобновляемы. Поэтому на них следует смотреть (и управлять ими) несколько иначе, чем на остальные ресурсы. Описанные рекомендации помогут организациям извлечь из них наибольшую выгоду.

Литература к главе 4

- DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

- Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Глава 5. От стратегического ресурса к ценнейшему активу

5.1. Повестка дня для совета директоров: развитие концепции управления информационными ресурсами

В 1995 году агентство Reuters опубликовало широко обсуждавшийся отчет «Информация как актив: Невидимая золотая жила»¹⁶², основанный на интервью с 500 менеджерами. Из них 25 % считали информацию самым важным активом, однако более 40 % заявили, что их компании еще не осознали ценность своей информации. Авторы отчета заключили: «Цифровая революция уже здесь – немногие активы так важны, как точная и своевременная информация. В завтрашнем информационном обществе информация станет наиболее доминирующим товаром». Они задались вопросами: «Сколько компаний пожинают плоды своего информационного преимущества? Оно имеет реальную власть и реальную

162 Reuters. Information as an Asset: The Invisible Goldmine. London: Reuters, 1995. _ URL: https://journals.sagepub.com/doi/pdf/10.1177/0266382954235537?casa_token=-5vYiWTEt1EAAAAA:hooT22bPYs5KTAzKBsFnokBDT38D_yZcwnXcaHLzI_-6aIMMXKuHN8PoVsVCbRBdBU2jRD6JtYYhsc.

ценность. У скольких компаний есть политики, которые позволяют контролировать, защищать, повышать ценность данных и управлять ими? Информация должна быть количественно измерена и оценена, а также эффективно управляться и использоваться».

Таким образом, на повестку дня организаций был вынесен вопрос о переходе на работу с данными как корпоративным активом.

Часто в качестве важного этапного документа, связанного с формированием представления о данных как об активе, упоминается вышедший почти одновременно с отчетом Reuters отчет комитета Хоули (The Hawley Committee) «Информация как актив: повестка дня для совета директоров»^{163, 164, 165, 166}. Документ был подготовлен под эгидой аудиторской компании KPMG (в рамках программы IMPACT¹⁶⁷) специально созданным комитетом, который возглавлял Роберт Хоули (Robert Hawley) – в то время генеральный директор британской компании Nuclear Electric. Комитет состоял из высокопоставленных менеджеров, руководивших компаниями из различных секторов. Участники обсуждали два вопроса.

1. Располагали ли их организации информацией, которая являлась стратегическим активом?
2. Понимали ли они и управляли ли этими активами так же, как понимали другие свои стратегические активы, используя и защищая их должным образом?

Комитет привлек к дискуссиям по указанным вопросам менеджеров высшего звена более 60 государственных и частных организаций.

Участники рассмотрели ситуацию с основными видами информационных ресурсов, идентифицируемых в большинстве привлеченных организаций. В качестве таких ресурсов были выделены^{168, 169}:

- информация о рынке и покупателях;
- информация о продуктах;
- специальные знания и информация, необходимая для работы в определенной области (этот вид сведений в настоящее время частично поддерживается с помощью методов управления знаниями, но на момент подготовки доклада Хоули управление знаниями не было устоявшейся концепцией);
- информация о бизнесе (экономическая, политическая и другая информация, лежащая в основе

¹⁶³ Oppenheim C., Stenson J., Wilson R. M. S. The attributes of information as an asset // New Library World, 2001, 102 (11): 458–464. DOI: 10.1108/EUM0000000006203 – URL: <https://www.libqual.org/documents/admin/stenson.pdf>.

¹⁶⁴ The Hawley Committee. Information as an Asset: The Board Agenda. London: KPMG, 1995. – URL: <https://www.cilip.org.uk/page/informationasset>.

¹⁶⁵ Ward S., Carter D. Information as an asset – Today's Board Agenda: The value of rediscovering gold // Business Information Review, 2019, 36 (2): 53–59. DOI: 10.1177/0266382119844639 – URL: https://www.researchgate.net/publication/332763457_Information_as_an_asset_-_Today's_Board_Agenda_The_value_of_rediscovering_gold.

¹⁶⁶ Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

¹⁶⁷ KPMG IMPACT – программа и платформа для поддержки и расширения возможностей специалистов KPMG по оказанию помощи клиентам в достижении их целей (<https://home.kpmg/xx/en/home/insights/2020/06/kpmg-impact.html>).

¹⁶⁸ Oppenheim C., Stenson J., Wilson R. M. S. (2003⁶). Studies on information as an Asset II: Repertory grid // Journal of Information Science, 2003, 29 (5): 419–432. DOI: 10.1177/01655515030295007 – URL: https://repository.lboro.ac.uk/articles/journal_contribution/Studies_on_Information_as_an_Asset_II_Repertory_Grid/9413912.

¹⁶⁹ Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

функционирования бизнеса, например цены акций);

- управленческая информация (особенно та, на которой базируются основные решения в области политики, конкурентных преимуществ или стратегических планов, например экономическая статистика или информация о затратах);

- информация о человеческих ресурсах;

- информация о поставщиках;

- отчетная информация (информация, требуемая различными регулирующими органами).

Обсуждение подтвердило: несмотря на то что информация – чрезвычайно ценный актив, ее управление редко рассматривается на уровне совета директоров. Как следствие, организации упускали возможности и подвергались рискам.

Было отмечено, что, хотя нематериальные активы (например, бренды и интеллектуальная собственность) обсуждаются в деловой литературе и связанные с ними вопросы доводятся до сведения советов директоров, информация по-прежнему остается загадочным и малообсуждаемым ресурсом. Она, как правило, попадала в поле зрения высшего руководства, только когда случались происшествия с катастрофическими последствиями.

Комитет Хоули предложил, чтобы вся значимая информация в организации, независимо от ее назначения, была должным образом идентифицирована (не обязательно с целью бухгалтерского учета) для рассмотрения в качестве бизнес-актива. Совет директоров должен выполнять соответствующие обязанности в отношении информационных активов таким же образом, как и в отношении других активов, например имущества. Это подразумевает новый подход к обращению с информацией, и нужно, чтобы совет директоров четко прояснил для менеджмента организации, какие действия следует предпринимать и кто несет ответственность за них. При этом информация должна рассматриваться как кросс-функциональный актив, а не принадлежащий конкретным бизнес-функциям.

Комитет выработал набор рекомендаций для советов директоров с целью повышения значимости как самой информации, так и практики управления ею на самом высоком уровне. Рекомендации сформулированы в виде десяти пунктов повестки дня. Первые три пункта касаются выстраивания работы с информацией на уровне совета, остальные посвящены политике работы с информацией на уровне организации в целом.

Прежде всего совет директоров должен убедиться в том, что его деятельность осуществляется таким образом, что:

- 1) информация, которую он использует, необходима и достаточна для его целей;

- 2) он осведомлен и правильно сориентирован относительно информационных аспектов всех вопросов, включенных в повестку дня;

- 3) практика использования им информации как на коллективном, так и на индивидуальном уровне соответствует применяемым законам, нормативным актам и этическим стандартам.

Совет также должен выработать политику организации в отношении информационных активов и определить, как будет анализироваться и оцениваться соблюдение этой политики, включая следующие аспекты:

- 4) идентификацию информационных активов и разделение (классификацию) их на те, что заслуживают особого внимания, и те, что к ним не относятся;

- 5) контроль за тем, что имеющееся количество и качество информации обеспечивают эффективную работу, гарантируя, что на каждом уровне предоставляемая информация необходима, достаточна, своевременна, надежна и согласованна;

- 6) надлежащее использование информации в соответствии с применяемыми правовыми и нормативными требованиями, операционными и этическими стандартами, а также ролями и обязанностями по созданию, хранению, доступу, изменению и уничтожению данных;

- 7) способность, соответствие и подготовку персонала для решения задач по защите и развитию информационных активов;

- 8) защиту информации от кражи, потери, несанкционированного доступа, злоупотреблений и некорректного использования, включая информацию, которая является собственностью других лиц;

- 9) обслуживание информационных активов и их надлежащее использование с максимальной выгодой

для организации, включая юридическую защиту, лицензирование, повторное использование, объединение, повторное предоставление, публикацию и уничтожение;

10) стратегию в отношении информационных систем и реализацию этой стратегии с уделением особого внимания возникающим затратам, выгодам и рискам.

К повестке дня был приложен контрольный перечень (чек-лист) и пояснительные примечания, описывающие мероприятия, необходимые для ее выполнения¹⁷⁰.

Основная цель комитета состояла в том, чтобы советы директоров взяли информацию под контроль. Они также рекомендовали определить информацию как организационный актив, поскольку каждый совет директоров имеет дело с управлением активами и отчетностью. Это продвинуло восприятие информации как ресурса в рамках концепции управления информационными ресурсами (IRM) в сторону восприятия ее как актива со своим собственным правовым положением в организации^{171, 172}.

5.2. Данные как актив

Как уже отмечалось в главе 4, концепция управления информационными ресурсами (IRM) сфокусирована на увеличении эффекта от данных прежде всего с точки зрения оправдания затрат на их сопровождение. Меньший акцент делается на вероятных будущих выгодах от их использования. Такой подход ограничивает интерпретацию данных как динамичного и изменяющегося ресурса, который способен приобретать и терять ценность в зависимости от контекста и применения¹⁷³. Поэтому следующим логичным шагом формирования организациями рационального отношения к своим информационным ресурсам стал переход к представлению о данных как корпоративном активе.

Под активом (asset) понимается имеющийся в собственности или контролируемый экономический ресурс, содержащий в себе или производящий ценность. При этом он может быть конвертирован в деньги (монетизирован)¹⁷⁴.

Международные стандарты финансовой отчетности (МСФО) определяют следующие важнейшие характеристики нематериального актива^{175, 176}:

¹⁷⁰ The Hawley Committee. Information as an Asset: The Board Agenda. London: KPMG, 1995. – URL: <https://www.cilip.org.uk/page/informationasset>.

¹⁷¹ Oppenheim C., Stenson J., Wilson R. M. S. The attributes of information as an asset // New Library World, 2001, 102 (11): 458–464. DOI: 10.1108/EUM000000006203 – URL: <https://www.libqual.org/documents/admin/stenson.pdf>.

¹⁷² Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

¹⁷³ Oppenheim C., Stenson J., Wilson R. M. S. The attributes of information as an asset // New Library World, 2001, 102 (11): 458–464. DOI: 10.1108/EUM000000006203 – URL: <https://www.libqual.org/documents/admin/stenson.pdf>.

¹⁷⁴ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

¹⁷⁵ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

¹⁷⁶ Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

- отсутствие физической формы;
- идентифицируемость (возможность разделения, продажи, передачи, лицензирования, аренды или обмена);
- контроль со стороны организации;
- наличие вероятной будущей экономической выгоды.

Если рассмотреть более подробно основные характеристики любого актива (материального и нематериального), можно убедиться, что данные и информация обладают ими в полной мере¹⁷⁷.

• Актив имеет потенциал обслуживания или будущие экономические выгоды: с точки зрения бухгалтерского учета какой-либо объект – актив, только если ожидается, что он обеспечит предоставление будущих услуг или экономические выгоды. Выгоды могут возникнуть как от использования, так и от продажи активов. Данные удовлетворяют этому требованию, поскольку обеспечивают возможность предоставлять услуги и принимать эффективные решения.

• Актив контролируется организацией: «контроль» тут означает ее способность извлекать выгоду из актива и, кроме того, запрещать или регулировать доступ других лиц к этой выгоде. Данные также подходят под это требование: если у организации есть данные, она одна имеет к ним доступ, если только она не продает или не предоставляет доступ другой стороне.

• Актив служит результатом прошлых транзакций: это означает, что контроль над активом был получен в ходе уже проведенных операций, таких как покупки, внутренняя разработка или выявление. Данные также удовлетворяют этому требованию. Они обычно собираются как побочный продукт транзакций (внутренняя разработка), могут быть результатом покупки (например, специализированной базы данных) или выявления новых полезных данных (например, путем анализа).

Данные удовлетворяют определению актива гораздо лучше, чем сотрудники или клиенты, которые также обычно рассматриваются как активы. Сотрудники и клиенты приводят к потенциальным услугам и будущим экономическим выгодам, но не являются собственностью компании: если сотрудники уйдут в отставку или клиенты сменят поставщиков, компания потеряет преимущества без компенсации. Данные нематериальны, однако именно потенциал обслуживания и экономические выгоды, а не физическая форма объекта имеют значение при оценке того, относится ли он к активам¹⁷⁸.

Основываясь на сходстве информационных активов с финансовыми, Джон Лэдли предложил «Общепринятые принципы учета информационных активов» (по аналогии с «Общепринятыми принципами бухгалтерского учета» [Generally Accepted Accounting Principles, GAAP])^{179, 180} (табл. 5.1).

¹⁷⁷ Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

¹⁷⁸ Там же.

¹⁷⁹ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.).

¹⁸⁰ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

Таблица 5.1.

Принципы учета информационных активов*

Принцип	Описание
Принцип подотчетности	Организация должна определить круг лиц, которые несут ответственность (и подотчетны) за данные и контент всех видов
Принцип управления данными как активом	Данные и контент всех видов являются активами и обладают всеми основными характеристиками других активов. Организация должна осуществлять управление ими, а также обеспечивать их защиту и учет наравне с прочими материальными и финансовыми активами
Принцип аудита	Правильность данных и контента подлежит периодическому аудиту (проводится независимой организацией)
Принцип должной осмотрительности	Если стало известно о риске, о нем следует доложить. Если риск возможен, это следует подтвердить. Риски в области данных включают риски, обусловленные неудовлетворительными практиками управления данными
Принцип непрерывности деятельности	Данные и контент жизненно важны для успешного и непрерывного осуществления бизнес-операций и управления (они не рассматриваются в качестве временного средства для достижения результатов или побочного продукта бизнеса)
Принцип разумного уровня оценки	Данные как актив следует оценивать на уровне, который представляется наиболее разумным, либо определяется самым простым из возможных способов
Принцип финансовой ответственности	Нарушения существующих юридических и этических норм при обращении с данными и контентом (злоупотребление или неудовлетворительное управление) влекут за собой финансовую ответственность
Принцип качества	Правильно передаваемый смысл, точность и надлежащая поддержка жизненного цикла данных и контента влияют на финансовое положение организации
Принцип риска	Владение данными и контентом сопряжено с риском, который нужно учитывать, в том числе и формально, посредством выделения средств либо на его страхование, либо на покрытие издержек от потенциальных последствий его материализации, либо на принятие мер по его минимизации
Принцип ценности	Данные и контент обладают ценностью, зависящей от способов их использования для достижения целей организации, от присущих им свойств товара и/или от их вклада в оценку деловой репутации организации. Ценность информации для организации снижается за счет затрат на ее обслуживание и перемещение

* Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

В 1999 году австралийские специалисты Дэниел Моуди (Daniel Moody) и Питер Уолш (Peter Walsh)

опубликовали статью «Измерение ценности информации: подход к оценке активов»¹⁸¹¹⁸². В ней отмечается, что многочисленные исследования, проведенные с привлечением широкого круга компаний, показали незначительную, если вообще какую-либо, положительную связь между инвестициями в ИТ и общими финансовыми показателями предприятия.

Чем отличаются организации, которые успешно используют ИТ для получения конкурентных преимуществ и достигают гораздо большего, чем их конкуренты, тратя при этом меньше? Исследования показывают, что общий фактор в подобных компаниях – то, что они фокусируются на самой информации, а не на технологии в качестве основы для достижения конкурентного преимущества.

В качестве примера Моуди и Уолш приводят результаты анализа деятельности четырех банков, которые потратили несколько лет на разработку систем поддержки своего основного бизнеса. Суммы, потраченные на это, различались в пять раз. Только один из них получил конкурентное преимущество – тот, что потратил наименьшую сумму.

Фактором, давшим успешному банку преимущество, была интеграция информации о клиентах. Технология – только средство доставки информации, а информация – это актив, который может быть использован для получения стратегического преимущества.

Аппаратное и программное обеспечение – всего лишь механизмы, используемые для создания и сопровождения информации; информация – базовый бизнес-актив, который производится и поддерживается с помощью этой технологии. Информация обеспечивает возможность предоставлять услуги, принимать эффективные решения, повышать производительность, достигать конкурентных преимуществ, а также может продаваться непосредственно как самостоятельный продукт. Если использовать производственную аналогию (рис. 5.1):

- данные – это сырье;
- техническое и программное обеспечение – это основные средства (здания и сооружения, а также машины и оборудование);
- информация – это конечный продукт, который доставляется клиенту.

Подавляющая часть расходов и выгод, связанных с использованием информационной системы, заключается в хранящейся информации, а не в аппаратном и программном обеспечении, используемом для ее хранения. Например, для работы информационной системы для клиентов в одной организации требовалось оборудование стоимостью \$500 000. Разработка собственного программного обеспечения обошлась в \$2 млн. Однако каждый год более \$10 млн тратилось на ввод информации о клиентах и поддержание ее в актуальном состоянии – эти расходы были скрыты в бюджетах зарплат отделов пользователей¹⁸³.

Для повышения эффективности стратегии организации в области ИТ должны быть сосредоточены на повышении и поддержании ценности информации (продукта), а не на системах и технологиях (производственном оборудовании). Бизнес-стратегии, как правило, оцениваются с точки зрения того, как они влияют на стоимость или прибыльность фирмы (результат). Аналогично, стратегии в области ИТ должны оцениваться с точки зрения того, как они способствуют достижению информационного результата – повышению ценности информации, предоставляемой пользователям.

¹⁸¹ Статья была представлена на 7-й Европейской конференции по информационным системам (Seventh European Conference on Information Systems – ECIS'99, Copenhagen Business School, Frederiksberg, Denmark, 23–25 June, 1999).

¹⁸² Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

¹⁸³ Там же.



Рис. 5.1.

Процесс производства информации*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

Несмотря на признание в качестве самостоятельного актива, стоимость информации до сих пор слабо поддается количественному измерению. Хотя для обеспечения сбора, хранения, обработки и обслуживания данных требуются огромные и постоянно растущие объемы организационных ресурсов, информационные активы, как правило, не получают финансового признания в балансе. В то время как аппаратные и (реже) программные активы капитализируются, оценка информации в значительной степени игнорируется, хотя это гораздо более ценный актив с точки зрения бизнеса.

Таким образом, для обеспечения по-настоящему эффективной работы с информацией как с активом необходима выработка определенных подходов к измерению ее стоимости.

Как и другие организационные активы, информация неразрывно связана с затратами (сколько стоит ее приобретение, хранение и обслуживание) и ценностью (какую выгоду она приносит организации). Однако на этом сходство заканчивается. Информация не подчиняется тем же законам экономики, что другие активы, – она обладает уникальными свойствами. Чтобы измерить ценность информации, необходимо эти свойства понять.

Моуди и Уолш попытались определить природу информации как актива, выделив ряд общих принципов или законов, которые регулируют ее поведение как экономического блага¹⁸⁴.

5.3. Семь законов информации

Первый закон. Информация может использоваться совместно (неограниченным количеством пользователей)

Вероятно, наиболее уникальная характеристика информации как актива – возможность ее одновременного использования неограниченным количеством людей, направлений бизнеса и организаций.

Эту характеристику мы уже рассматривали в главе 4, обсуждая свойства данных как ресурса. Рассматривая данные как актив, следует обратить внимание на то, что совместное использование осуществляется без потери ценности информации для каждой из заинтересованных сторон (рис. 5.2). Это сильно отличается от свойств других активов. Большинство активов – «присваиваемые». Их может использовать только кто-то один. Например, совместное использование финансов, оборудования или персонала несколькими бизнес-направлениями приводит к тому, что каждое из них получает только часть

¹⁸⁴ Там же.

ценности активов. Напротив, информация может предоставляться несколькими бизнес-направлениями, обеспечивая одинаковую ценность для всех сторон, как если бы она была в их исключительном пользовании. Таким образом, с точки зрения организации, ценность является суммируемой, а не разделяемой между пользователями.^[66]

Например, если один отдел в организации ведет таблицу курсов валют, эта информация может быть передана другому отделу без увеличения или снижения ценности информации для первого отдела. Альтернативой этому было бы принятие решений в отсутствие необходимой информации (с последующей потерей эффективности бизнеса) или ведение вторым отделом собственной таблицы курсов валют (с последующим увеличением затрат).

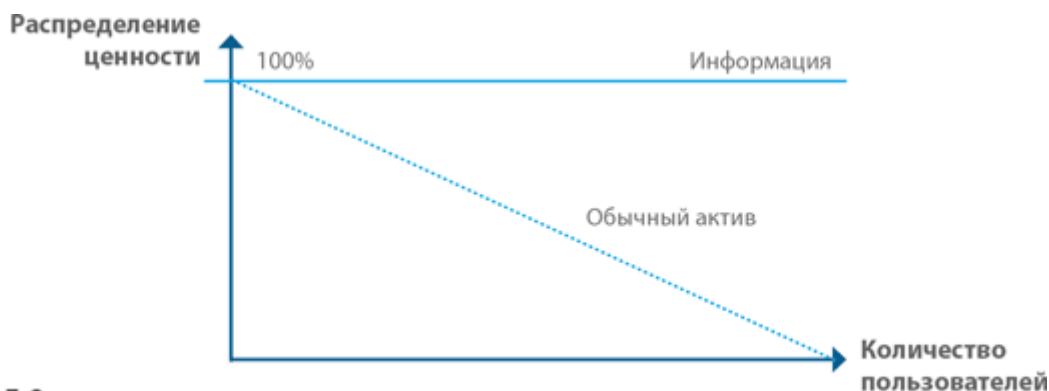


Рис. 5.2.

Совместное использование информации*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

Одна из основных проблем совместного использования на практике заключается в том, что обмену информацией препятствуют барьеры как институционального, так и личного характера. Одностороннее понимание лозунга «Знание – сила» приводит к тому, что люди не всегда охотно делятся знаниями. Придерживание информации – когда одни сотрудники отказывают другим в доступе к имеющимся у них сведениям – представляет собой потерю возможностей для бизнеса, поскольку это препятствует реализации потенциальной ценности.

Информацию можно не только неограниченно совместно использовать, но и неограниченно копировать (это свойство тоже рассматривалось в главе 4). Часто копирование – результат необоснованного отказа от совместного использования. Дублирование информации не удваивает ее ценность – две копии имеют то же значение, что и одна, потому что новая информация не появляется. Без повышения ценности дублирование приводит к дополнительным затратам: на повторную обработку данных в нескольких системах, на хранение, дополнительные усилия по разработке систем, взаимодействие для обеспечения согласованности данных, на усилия по ручной сверке.

Ряд исследований показал, что в средних и крупных организациях, как правило, существует порядка десяти копий каждого элемента корпоративных данных. Очевидно, что затраты на поддержание такого уровня избыточности огромны. То, что это может происходить столь явно неконтролируемым образом, отражает отсутствие практики применения финансовой отчетности в отношении деятельности по управлению информацией.

Второй закон. Ценность информации возрастает с увеличением использования

Для большинства активов характерно снижение отдачи от использования – чем больше они используются, тем меньше их ценность. Например, транспортные средства амортизируются в зависимости от пройденных километров, воздушные суда – в зависимости от часов полета, оборудование – в зависимости от часов эксплуатации. Но ценность информации возрастает по мере ее использования, с расширением применения она демонстрирует все большую отдачу. Основные затраты, связанные с информацией, приходятся на ее сбор, хранение и обслуживание – дополнительные затраты при ее использовании не так велики (рис. 5.3).

Информация сама по себе не имеет реальной ценности – она становится ценной, только когда люди ее используют. Если информация не используется, она не может привести к экономическим выгодам и, следовательно, не является активом. Неиспользованная информация должна быть сопряжена с ответственностью, поскольку из нее не извлекается никакой ценности и организация несет неоправданные затраты на ее хранение и обслуживание. Во многих организациях существует большое количество информации, которая собирается и хранится, но никогда не используется, представляя собой некую разновидность отходов¹⁸⁵.

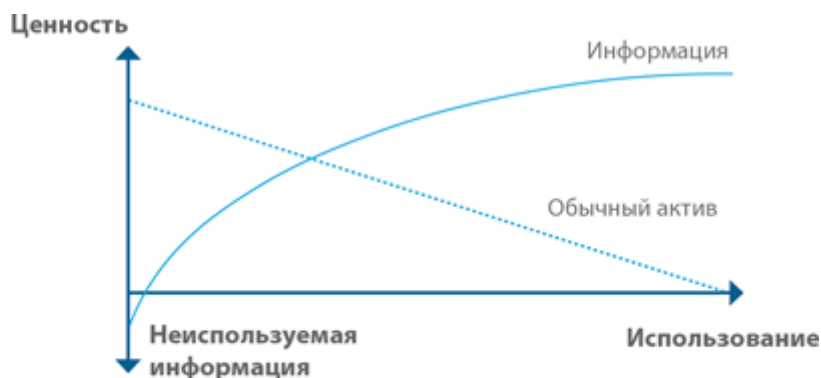


Рис. 5.3.

Повышение ценности информации с увеличением использования*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

Необходимыми условиями для эффективного использования информации являются:

- знание о ее существовании;
- знание того, где она находится;
- наличие к ней доступа;
- знание того, как ее использовать.

Информация достигает своего наивысшего потенциала, когда все в организации знают, где она находится, имеют к ней доступ и понимают, как ее использовать. Она имеет самый низкий потенциал, если люди даже не знают, что она есть. В большинстве организаций существует огромное количество цифровой информации, которая может быть использована в интересах бизнеса. При этом возможности часто не реализуются, потому что сотрудники не знают о ее существовании. Каким бы странным это ни казалось, лишь немногие организации имеют каталог своих информационных активов и мест их расположения – реестр информационных активов. Очевидно, что такая ситуация была бы недопустима для физических или финансовых активов. Однако информация не подпадает под те же требования к аудиту, что и другие активы¹⁸⁶.

Еще один важный вопрос – грамотность в области данных (data literacy). Часто предполагается, что, если людям будет предоставлена информация, процесс принятия решений автоматически улучшится. Однако качество принятия решений зависит как от качества предоставляемой информации, так и от способности лиц, принимающих решения, интерпретировать и использовать ее. Нет смысла повышать точность и своевременность предоставления данных, если сотрудники не знают, как их эффективно использовать. Поэтому наряду с повышением качества самой информации необходимо уделять не меньшее внимание повышению уровня грамотности в области данных.

¹⁸⁵ В этой связи в главе 6 мы рассмотрим понятие data ROT («информационная гниль»).

¹⁸⁶ В главе 4 мы уже приводили создание реестра наборов данных в качестве одной из главных рекомендаций компаниям, реализующим концепцию управления информационными ресурсами.

Третий закон. Пригодность информации со временем снижается

В главе 4, обсуждая свойства данных как ресурса, мы уже рассматривали такую их характеристику, как обесцениваемость. Остановимся на ней подробнее в контексте понятия срока годности.

Как и большинство других активов, информация имеет тенденцию обесцениваться со временем. Скорость, с которой теряется ценность, зависит от вида информации. Например, как только клиент изменил адрес, старый адрес может немного значить или не значить ничего. С другой стороны, показатели продажи товаров могут быть актуальны в течение многих лет в будущем. Как показано на рисунке 5.4, информация имеет три срока годности: операционный, нормативный (определенный нормативно-правовыми актами) и срок годности для поддержки принятия решений.

Например, реквизиты сезонного железнодорожного билета действительны только в течение сезона – это операционный срок годности. Однако для целей принятия решений (определения тенденций и закономерностей) может иметь значение информация о продажах билетов, например за последние пять лет, – это срок годности для принятия решений. Кроме того, в отношении отдельных видов информации существуют требования нормативно-правовых актов по хранению записей в течение десятков лет – нормативный срок годности.

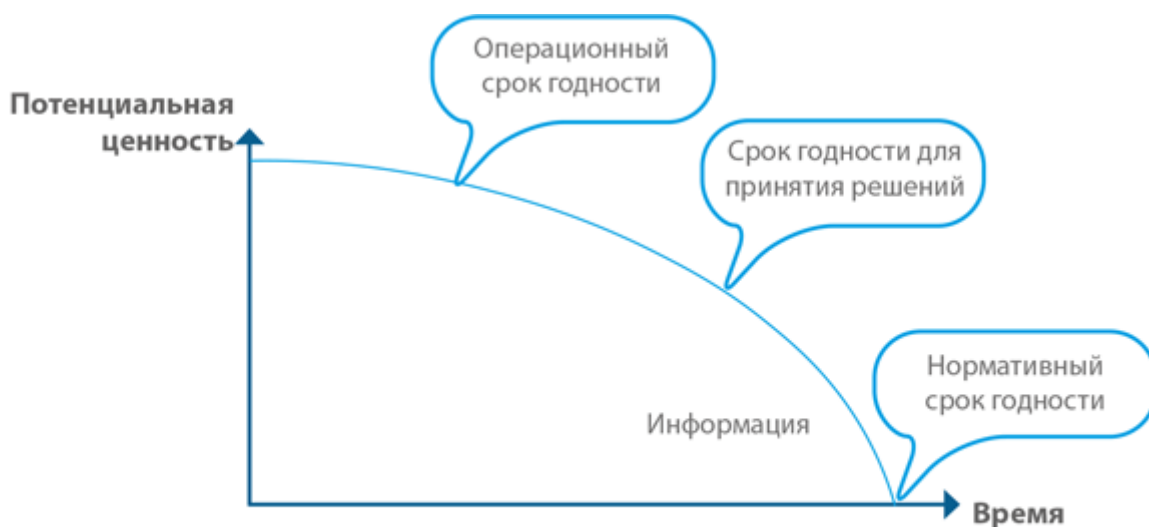


Рис. 5.4.

Обесценивание информации с течением времени*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

На операционном уровне информация имеет относительно короткий срок годности. Здесь в основном обрабатываются наиболее свежие данные, например текущие адреса клиентов или их последние номера счетов. Информационные системы, поддерживающие операционную деятельность, часто перестают использовать информацию и предоставлять к ней доступ, как только время ее существования превысит операционный срок годности. Срок годности для принятия решений гораздо более длительный. Чтобы сделать информацию доступной на протяжении этого срока, организуются специальные хранилища (Data Warehouse, DW), которые поддерживают работу с историческими данными.

Четвертый закон. Ценность информации повышается с увеличением точности

Чем точнее информация, тем она ценнее. Неточная информация может дорого стоить организации из-за операционных ошибок или неправильного принятия решений. Необходимый уровень точности сильно зависит от вида информации и того, как она используется. Для некоторых сведений может требоваться стопроцентная точность, в то время как для других сведений точности на уровне 80 % может быть достаточно.

Как показано на рисунке 5.5, существует область снижения отдачи, после которой дальнейшее повышение точности дает мало дополнительных преимуществ – стопроцентно точная информация редко требуется в контексте бизнеса. Также нужно учитывать, что, как только точность информации падает ниже

определенного уровня, она становится пассивом, а не активом. В этот момент она превращается в дезинформацию и ее перестают использовать.

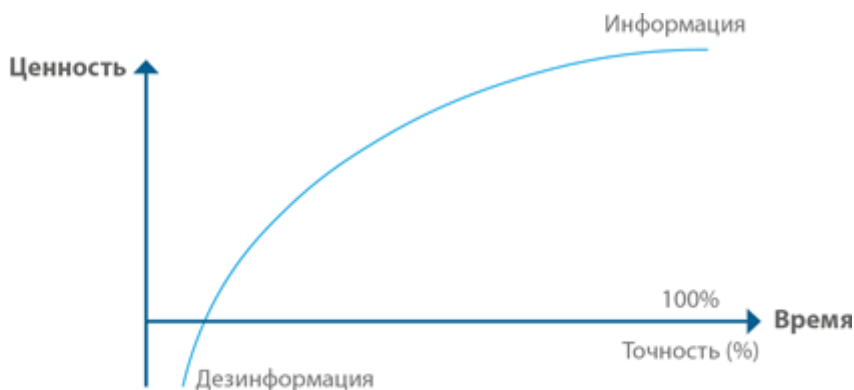


Рис. 5.5.

Повышение ценности информации с увеличением точности*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

Пятый закон. Ценность информации повышается при объединении с другой информацией

Информация, как правило, становится более ценной, когда ее можно сравнить и объединить с другой информацией. Например, информация о клиентах и информация о продажах сами по себе представляют ценность. Однако с точки зрения бизнеса возможность связать эти два набора данных вместе принесет большую выгоду. Способность соотносить характеристики покупателей с моделями покупок может помочь нацелить маркетинговые усилия на то, чтобы правильные продукты продвигались нужным людям в нужное время.

При подготовке информации для анализа и принятия решений, как правило, требуется консолидация данных из широкого спектра различных систем, поддерживающих операционную деятельность. В большинстве организаций отсутствие интеграции данных – серьезное препятствие при осуществлении аналитической деятельности. Данные из разных систем часто могут быть объединены только при значительном ручном вмешательстве и устранении расхождений. Львиная доля усилий по созданию корпоративных информационных систем приходится на извлечение и согласование данных из различных источников и преобразование их в общие форматы¹⁸⁷.

Большинство преимуществ интеграции может быть достигнуто за счет стандартизации относительно небольшого процента данных. Наиболее важными для целей интеграции элементами данных являются идентификаторы (позволяющие связывать информацию из разных источников об одном объекте воедино) и системы кодирования (которые используются в качестве основы для агрегирования данных в управленческой отчетности). В соответствии с законом Парето (или правилом 80/20), интеграция 20 % данных обычно дает 80 % преимуществ. Дальнейшая интеграция обычно сопровождается снижением отдачи (рис. 5.6).

Шестой закон. Увеличение объема информации не всегда приводит к повышению ценности

В большинстве случаев чем больше у вас определенного ресурса (например, финансов), тем лучше для вас. Общая проблема управления – решение вопроса о том, как распределить ограниченные ресурсы между конкурирующими заинтересованными сторонами. Однако с ростом использования информационных технологий говорить о дефиците информации приходится все реже. Фактически, самая большая проблема для большинства организаций сегодня состоит не в недостатке информации, а в ее переизбытке. Исследования в области психологии показывают, что у людей способность обрабатывать информацию строго ограничена, а эффективность принятия решений снижается, как только объем

¹⁸⁷ Более подробно вопросы интеграции обсуждаются в главе 12 «Обеспечение доступности и обслуживание данных: основы».

информации превышает определенную оптимальную величину (рис. 5.7).

Здесь существует интересный парадокс. Известный экономический закон убывающей предельной полезности гласит, что чем больше у людей определенного товара, тем ниже для них его ценность. Однако экспериментальные исследования показывают, что ощущение получения ценности у лиц, принимающих решения, продолжает расти даже за пределами точки перегрузки. Люди, принимающие решения, как правило, ищут больше информации, чем они могут обработать. Хотя избыток информации приводит к снижению производительности, он вселяет уверенность и повышает удовлетворенность лица, принимающего решение. Таким образом, большинство людей не подозревают о собственных ограничениях по обработке информации и считают, что чем больше информации, тем лучше.

Седьмой закон. Информация не истощается

В главе 4 мы уже рассматривали по отдельности свойства нерасходуемости и возобновляемости данных. Поговорим немного об их сочетании.

Большинство ресурсов истощаются – чем больше вы их используете, тем меньше у вас остается. Однако информация самогенерируется – чем больше вы ее используете, тем у вас ее больше. Это связано с тем, что новая информация часто создается в результате обобщения, анализа или объединения различных источников данных. Исходная информация остается, а производная добавляется к существующей базе активов. По сути, именно поэтому информация не дефицитный ресурс. Такие методы, как интеллектуальный анализ данных (data mining), используются специально для получения новой информации из существующих хранилищ данных.

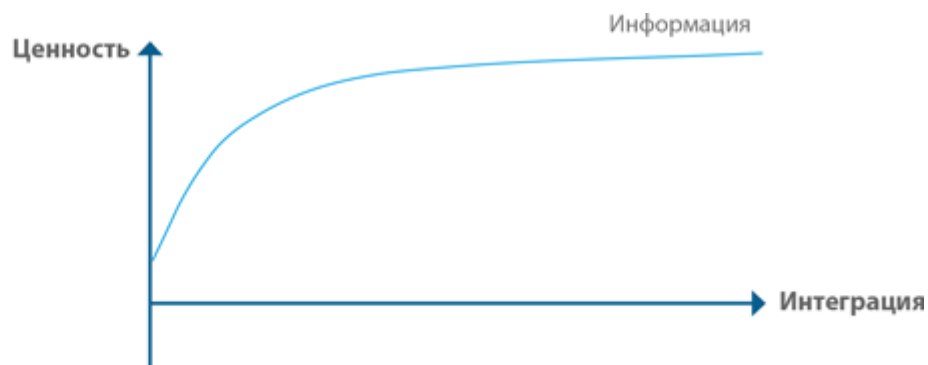


Рис. 5.6.

Увеличение ценности информации в процессе интеграции*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

По сравнению с другими основными видами корпоративных активов (люди, финансы, имущество) управление информацией осуществляется наименее эффективно. Если бы большинство организаций управляли своими финансами так же плохо, как и информацией, они быстро завершили бы свою деятельность. Большие объемы дублирования, отсутствие стандартизации и недостаточное внимание к качеству информации редко допускается в отношении других активов. Сложно однозначно объяснить, почему так происходит. Поэтому определение законов информации может послужить основой для совершенствования управления. Чтобы эффективно управлять информацией, необходимо понимать ее законы, так же как необходимо понимать законы человеческого поведения, чтобы эффективно управлять людьми, или законы экономики, чтобы эффективно управлять финансами.

Как уже отмечалось в конце предыдущего раздела, важное условие управления данными как активом – способность организации осуществлять их оценку. Вопросам определения ценности данных пока уделяется мало внимания. Однако они могут оказать огромное влияние на ИТ-индустрию в будущем. Доля ресурсов, выделяемых организациями на обработку информации, растет экспоненциально, но слабо отражается в финансовой отчетности. Одна из проблем заключается в том, что в настоящее время не

существует общепринятого подхода к измерению ценности информационных активов^{188, 189, 190, 191, 192}.



Рис. 5.7.

Соотношение объема информации и ее ценности*

* Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

Ценность информации с трудом поддается количественной оценке, поскольку сильно зависит от контекста и использования. Кроме того, потенциальная ценность информационного актива – ненадежный показатель его фактической стоимости. Если такая ценность никогда не будет реализована, организация не получит никакой выгоды. Поэтому при оценке информации часто используется подход на основе определения стоимости утраты (deprival value), т. е. стоимость информационного актива для организации, пытаются определить как стоимость потерь, которые она понесла бы, лишившись этого актива. Правила бухгалтерского учета не допускают включения таких нематериальных активов в бухгалтерский баланс, даже если они являются основным источником стоимости в бизнесе. Таким образом, бухгалтерский учет пока отстает от экономики знаний. Но то, что нематериальные активы не могут быть учтены в балансе,

¹⁸⁸ Collins V., Lanz J. Managing Data as an Asset // The CPA Journal, June 2019. – URL: <https://www.cpajournal.com/2019/06/24/managing-data-as-an-asse>.

¹⁸⁹ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

¹⁹⁰ Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

¹⁹¹ Oppenheim C., Stenson J., Wilson R. M. S. The attributes of information as an asset // New Library World, 2001, 102 (11): 458–464. DOI: 10.1108/EUM0000000006203 – URL: <https://www.libqual.org/documents/admin/stenson.pdf>.

¹⁹² Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

вовсе не означает, что они не учитываются и не должны учитываться^{193, 194}.

В следующем разделе мы рассмотрим ряд хорошо проработанных на сегодня моделей оценки информационных активов.

5.4. Модели оценки информационных активов

Даглас Лейни в книге «Инфономика: информация как актив: монетизация, оценка, управление»¹⁹⁵ приводит перекликающийся с описанными в предыдущем разделе семью законами перечень уникальных характеристик информационных активов.

- Высокая степень повторного использования.
- Способность превращения в ликвидный актив в широком спектре контекстов.
- В настоящее время данные не подлежат бухгалтерскому учету, поэтому организации в большинстве случаев не обязаны раскрывать сведения о своих информационных активах.

- При обмене информация не облагается налогом.
- Легкость тиражирования.
- Несложная и мгновенная передача.
- Широкие перспективы по предоставлению новых выгод.

Все эти характеристики открывают неограниченные возможности для монетизации данных. В этой связи задача определения подходов к измерению их ценности становится еще более актуальной.

Лейни описывает несколько моделей оценки информационных активов, разработанных в рамках исследований, проведенных компанией Gartner. Состав моделей отражен на рисунке 5.8.

¹⁹³ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

¹⁹⁴ Evans N., Price J. Enterprise information asset management: the roles and responsibilities of executive boards // Knowledge Management Research & Practice, 2015. 14: 353–361. – URL: <https://www.experiencematters.com.au/wp-content/uploads/2018/03/Enterprise-information-asset-management-the-roles-and-responsibilities-of-executive-boards.pdf>.

¹⁹⁵ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)



СЕМЬ ЗАКОНОВ ИНФОРМАЦИИ

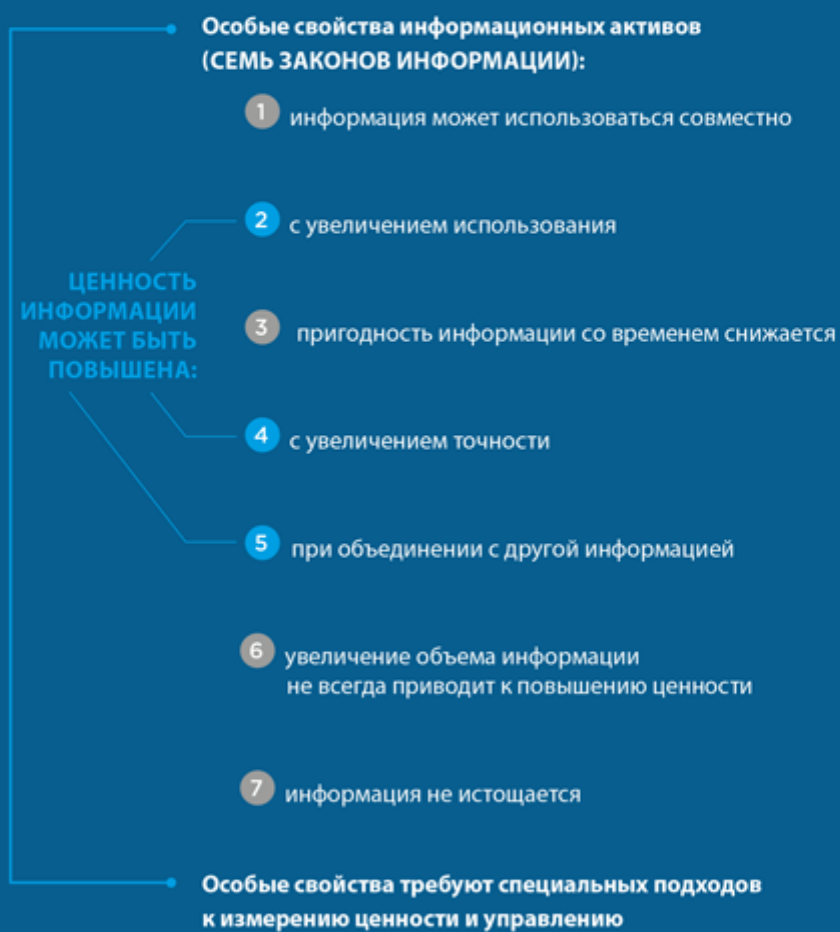




Рис. 5.8.

Модели Gartner для оценки информационных активов*

* Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Модели являются результатом многолетнего сотрудничества с экспертами по оценке, бухгалтерами, экономистами и клиентами Gartner. Они отражают как основные, так и финансовые подходы к расчету стоимости.

Основные (базовые) модели учитывают аспекты информации, связанные с ее качеством или влиянием на показатели эффективности организации. Финансовые модели разработаны для измерения ценности информации в денежном выражении путем адаптации принятых методов оценки традиционных активов.

Следует заметить, что приводимые модели на данный момент не являются официально принятыми организациями, определяющими стандарты в области бухгалтерского учета и финансовой отчетности. Они предназначены исключительно для внутреннего использования компаниями и учреждениями при оценке и сравнении ценности их информационных активов.

Модели могут быть применимы к конкретным наборам данных, но больше преимуществ из них можно извлечь, если рассматривать логические группировки (портфели) связанных информационных активов. Например, данные о клиенте, о продукте, об обслуживании, данные контакт-центра или данные о сотрудниках.

Все приводимые методы, как и методы оценки любого вида активов, основаны на ряде допущений. Важно, чтобы эти допущения были правильно описаны и последовательно применялись.

Хотя дискретные оценки на определенный момент времени безусловно полезны, они могут быть не столь значимыми, как:

- понимание различий между реализованной, прогнозируемой и потенциальной ценностью информационного актива;
- отслеживание роста или снижения ценности информационного актива с течением времени;
- сопоставление или объединение измеряемых показателей с целью выработки новых перспективных

решений для тех или иных задач.

Выбор той или иной модели зависит от цели оценки. Некоторые из моделей определяют опережающие показатели, некоторые – запаздывающие, а некоторые могут использоваться для определения и тех и других¹⁹⁶.

Во многих случаях имеет смысл одновременно осуществлять измерения по нескольким разным моделям в интересах различных функциональных блоков организации (ИТ, финансы, отдельные направления бизнеса и т. п.). Возможно также использование различных сочетаний подходов¹⁹⁷.

1. Основные модели

Основные модели оценки ориентированы на организации или отдельные подразделения, которые еще не готовы или не имеют острой необходимости определять денежную стоимость своих информационных активов. Эти модели подходят для оценки качества информационных активов и сравнения их потенциальной и фактической полезности с целью улучшения управления данными организации. Они также могут быть применены для определения опережающих показателей потенциальной экономической выгоды от использования информационных активов.

Внутренняя ценность информации

Обзор

Показатель внутренней ценности информации (Intrinsic Value of Information, IVI) определяет связанную с ней предполагаемую выгоду. Он позволяет проводить укрупненные сравнения различных групп информационных активов независимо от того, каким образом они используются в настоящее время. Этот метод определяет, насколько верными и полными являются представленные в информационных активах данные и насколько вероятно, что такие данные имеются у других организаций. Предполагается, что более качественная и доступная информация, которая уникальна и контролируется с точки зрения принадлежности данной организации, обладает более высоким внутренним ценностным потенциалом.

Модель IVI применима для определения связанных с информацией инвестиционных приоритетов (в процессе рассмотрения различных источников данных или инициатив). Например, показатель IVI может быть особенно полезен для распределения усилий и инвестиций в области качества данных или информационной безопасности.

Формула расчета

$$IVI = \text{Validity} * \text{Completeness} * (1 - \text{Scarcity}) * \text{Lifecycle},$$

где:

Validity (валидность) – процент записей, признанных верными;

Completeness (полнота) – процентное соотношение текущего общего количества записей и потенциально возможного количества записей;

Scarcity (редкость) – процент представленных на рынке организаций (или конкурентов), которые, вероятно, имеют такие же данные;

Lifecycle (жизненный цикл) – разумный срок полезного использования любой единицы данных (записи), принадлежащей информационному активу (например, в месяцах).

Реализация

Эта модель хорошо подходит для использования ответственными сотрудниками с целью сравнения потенциальной полезности различных видов информации или контроля улучшения (или ухудшения) динамики потенциала отдельных информационных активов с течением времени. Оптимальное значение IVI – 1.0 (идеальная точность и полнота данных при отсутствии копий или версий какой-либо части этих данных, доступных за пределами организации). Для информации с высоким IVI, возможно, будет целесообразным расширение ее доступности и использования. При низком IVI может потребоваться

¹⁹⁶ Запаздывающие показатели отражают результаты деятельности организации за прошедший период. К опережающим относятся прогнозные показатели, помогающие предвидеть будущий результат (см., например, <https://www.lcbi.ru/company/news/356326/>).

¹⁹⁷ Там же.

удвоение усилий по управлению информационным активом и обеспечению его качества.

Преимущества и проблемы

Модель IVI – самая простая из используемых моделей оценки информации. Она поможет быстро сравнить ценностный потенциал различных информационных активов, определить качество данных, их конфиденциальность или проблемы с управлением активом. Кроме того, она позволяет выявить данные, которые, возможно, не следует хранить. Однако эта модель не учитывает применимость данных для достижения какой-либо конкретной бизнес-цели.

Бизнес-ценность информации

Обзор

Модель определения бизнес-ценности информации (Business Value of Information, BVI) рассматривает полезность информационного актива с точки зрения фактического применения в бизнесе (в отличие от модели IVI). В ней рассматривается, насколько полезна, актуальна информация и насколько она применима к бизнесу. Метод удобен для быстрого получения потенциальной пользы от информации в реальном мире. Например, при наличии конкурирующих бизнес-приоритетов модель можно использовать для согласования с ними приоритетов, связанных с информацией.

Формула расчета

$$BVI = \sum_{p=1}^n (Relevance_p) * Validity * Completeness * Timeliness,$$

где:

Relevance (релевантность) – насколько полезной данная информация может быть для одного или нескольких бизнес-процессов (от 0 до 1); n – количество бизнес-процессов или функций;

Validity (валидность) – процент записей, признанных верными;

Completeness (полнота) – процентное соотношение текущего общего количества записей и потенциально возможного количества записей;

Timeliness (своевременность) – насколько быстро новые или обновленные записи собираются и становятся доступными.

Реализация

Для реализации этого метода удобно иметь общее распределение бизнес-функций по всей организации. Уровень детализации распределения может различаться в зависимости от имеющихся возможностей по анализу функций и конкретных задач по оценке информации. С помощью измерения разрывов между фактической ценностью информации и ее потенциальной ценностью (в частности, с использованием оценок фактической и потенциальной релевантности) можно быстро определить возможности для лучшего использования информационных активов.

Преимущества и проблемы

Модель BVI связывает данные с фактической ценностью для бизнеса. Она относительно проста в реализации и может учитывать фактические и потенциальные сценарии. Этот подход полезен для выявления «темных данных» (неиспользуемых или малоиспользуемых)¹⁹⁸ и принятия решений об удалении ненужной информации. Однако определение коэффициентов релевантности данных бизнес-функциям может потребовать трудоемкого функционального анализа и быть весьма субъективным.

Ценность информации с точки зрения эффективности

Обзор

Модель определения ценности информации с точки зрения эффективности (Performance Value of Information, PVI) рассматривает фактическое (или ожидаемое) влияние информационного актива на достижение бизнес-целей, которые представлены в виде ключевых показателей эффективности (Key Performance Indicators, KPIs). Показатель PVI отвечает на вопрос, насколько наличие данной информации улучшает эффективность бизнеса. Для реализации модели требуется проведение контролируемого эксперимента, что позволяет осуществить итоговое, основанное на опыте измерение ценности

¹⁹⁸ Понятие «темные данные» будет рассмотрено в главе 6.

информационных активов.

Поскольку показатель PVI запаздывающий, он может быть менее полезным (чем IVI или BVI) для определения приоритетов инициатив в области данных или определения потенциальной ценности информации. Однако его использование наиболее предпочтительно для измерения реализованных бизнес-выгод в соответствии с установленными метриками. Он также применяется в качестве запаздывающего показателя для финансовой оценки информационного актива.

Формула расчета

PVI – это отношение, которое показывает улучшение ключевого показателя эффективности за счет задействования данного информационного актива. При этом оно экстраполировано на средний срок полезного использования записи.

Для нескольких ключевых показателей эффективности общее значение PVI может быть выражено как среднее значение отдельных PVI.

$$PVI = \left[\left(\frac{KPI_i}{KPI_c} \right) - 1 \right] * \frac{T}{t},$$

где:

KPI_i – экземпляры бизнес-процессов, использующие информационный актив (информированная группа);

KPI_c – экземпляры бизнес-процессов, не использующие информационный актив (контрольная группа);

T – средний срок полезного использования записи;

t – отрезок времени, в течение которой измерялся ключевой показатель эффективности.

Реализация

Использование модели PVI в идеале требует проведения контролируемого эксперимента, в котором определенные экземпляры бизнес-процесса используют информационный ресурс, не задействованный в других экземплярах. Это классический A/B-тест. Положительный PVI демонстрирует, что данные ценны для этого процесса; отрицательный PVI – что дополнительные данные снижают его эффективность. При определении PVI важно сохранять неизменными все другие аспекты улучшения процесса.

Преимущества и проблемы

PVI дает надежные, полученные опытным путем результаты измерений, которые являются хорошим средством прогнозирования или косвенным показателем при проведении финансовых оценок. С помощью данного подхода реализуются практические сценарии без необходимости анализа на уровне бизнес-функций. В то же время при использовании PVI требуется проведение одного или нескольких экспериментов, потенциально связанных с изменениями систем или процессов. Следует также заметить, что на результат влияет способ интеграции данных в процесс, причем модель не учитывает затраты на такую интеграцию.

2. Финансовые модели

Модели финансовой оценки информации полезны для тех организаций, которым необходимо определить, насколько эффективны информационные активы по сравнению с другими, какие средства нужно инвестировать в их сбор, управление, безопасность и развертывание и как отразить их стоимость при использовании в бизнес-операциях (например, в операциях слияния и поглощения, объединения обмена данными и т. п.). Эти экономические модели являются вариантами существующих моделей, которые используются экспертами по оценке и бухгалтерами при определении стоимости традиционных активов. В то же время они адаптированы с учетом особенностей, вытекающих из уникальных характеристик информации. В частности, с учетом того, что она не истощается, имеет высокую степень повторного использования и более ориентирована на лицензирование, чем на продажу.

Ценность информации с точки зрения затрат

Обзор

Модель определения ценности информации с точки зрения затрат (Cost Value of Information, CVI) рассматривает информационный актив через призму расходов, необходимых для генерации или сбора

образующих его данных. Дополнительно в ней может учитываться влияние на бизнес в случае, если информационный актив будет недоступен (например, поврежден или утерян) или украден (в частности, скопирован). Этот метод предпочтителен в случаях, когда актив не сильно востребован и его вклад в доход организации не может быть определен достаточно точно. Кроме того, эта модель может быть использована для оценки потенциального финансового риска повреждения, потери или кражи информационного актива.

Формула расчета

$$CVI = \frac{ProcExp * Attrib * T}{t} \left\{ + \sum_{p=0}^n Lost\ Revenue_p \right\},$$

где:

ProcExp – годовые затраты на процесс (процессы), связанный (связанные) со сбором данных;

Attrib – доля (процент) затрат на процесс, приходящаяся на сбор данных;

T – средняя продолжительность жизни записи;

t – период времени, в течение которого измеряются затраты на процесс;

Lost Revenue – доход, потерянный в результате повреждения, потери или кражи информационного актива;

n – количество периодов времени до повторного получения информации или до тех пор, пока потерянная или поврежденная информация больше не будет влиять на непрерывность бизнеса.

Реализация

Затраты на процесс и долю, относящуюся к сбору информации, может быть сложно определить, учитывая, что они могут быть собраны в ходе бизнес-операций, и в этом случае они обычно относятся на другие расходные статьи. Если бы была определена доля затрат на процесс, связанная с приобретением информационного актива, то эта сумма предположительно могла бы быть заявлена как стоимость актива, а не как затраты (несмотря на действующие правила бухгалтерского учета). Следует также учитывать стоимость репутационных или конкурентных рисков, если информация будет опубликована или украдена конкурентами.

Преимущества и проблемы

Модель CVI – лучшее средство оценки затрат на обновление информации и негативного влияния на бизнес в случае потери, кражи или повреждения. Бухгалтеры предпочитают этот метод как более консервативный и менее изменчивый подход к первоначальной оценке большинства нематериальных активов. Однако некоторые факторы требуют дополнительных подсчетов и субъективного взгляда. Эти затраты, скорее всего, уже учтены, поэтому CVI просто выражает ценность информации с точки зрения перевода ее из категории затрат.

Рыночная ценность информации

Обзор

Модель определения рыночной ценности информации (Market Value of Information, MVI) рассматривает потенциальную или фактическую финансовую стоимость информационного актива на открытом рынке. Как правило, монетизация данных осуществляется в результате передачи их торговым партнерам в обмен на деньги, товары, услуги или на других условиях, таких как льготные контракты. Тем не менее все чаще компании продают свои данные напрямую через информационные рыночные площадки или отраслевых информационных брокеров.

Метод MVI, как правило, неприменим для большинства видов информационных активов, если они не подлежат лицензированию или обмену. Однако по мере того как организации становятся все более опытными и активными в предоставлении своих данных на коммерческой основе, им следует применять эту модель.

Формула расчета

Специалисты Gartner модифицировали традиционный метод определения рыночной ценности исходя из того, что большая часть информации фактически не продается. Вместо этого она лицензируется. В этой связи принят во внимание фактор снижения конкурентоспособности информации, так как она становится все более распространенной на рынке. Поэтому в формулу добавлен переменный коэффициент дисконтирования (представленный в виде обратной премии – Premium), применяемый к

предположительной передаче права собственности (эксклюзивная цена – Exclusive Price, помноженная на количество лицензий – Number of Licensees) информационного актива:

$$MVI = \frac{\text{Exclusive Price} * \text{Number of Licensees}}{\text{Premium}}.$$

Реализация

MVI следует использовать при рассмотрении возможности монетизации информации посредством продажи или обмена. В идеале для определения эксклюзивной цены можно использовать модель CVI или EVI (описана далее). Эти методы помогут определить, сколько денег организация может потребовать за передачу полного права собственности на информационный актив другой организации. Затем следует определить или оценить, сколько вероятных сторон могут лицензировать эти данные в течение средней продолжительности жизни записи. Для определения числа вероятных лицензиаров применимы традиционные методы анализа рынка. Выяснить размер премии можно с помощью дополнительных опросов потенциальных лицензиаров.

Преимущества и проблемы

Модель MVI наиболее полезна для определения стоимости продаваемого или обмениваемого информационного актива. Она также применима для определения цены на информационный продукт или же может быть адаптирована для определения приемлемого размера лицензионных сборов за информационные продукты других организаций. Однако данный подход не особенно пригоден в отношении нерыночных информационных активов. Он включает в себя весьма субъективные факторы, которые могут потребовать тщательного рыночного анализа, например при определении или оценке эксклюзивной цены на информационный актив.

Экономическая ценность информации

Обзор

Модель определения экономической ценности информации (Economic Value of Information, EVI) позволяет рассчитывать чистую финансовую стоимость информационного актива, применяя для его оценки традиционный доходный подход, а затем вычитая затраты, связанные с жизненным циклом информации. Как и PVI, этот метод эмпирически вычисляет фактическую стоимость информационного актива. Это скорее запаздывающий показатель ценности, чем опережающий – если только первый срок получения дохода не может быть оценен достаточно точно.

Формула расчета

EVI учитывает реализованное изменение дохода, когда конкретный информационный актив задействован в одном или нескольких процессах, приносящих доход. Затем затраты на получение (AcqExp), администрирование (AdmExp) и применение (AppExp) данных будут вычтены.

$$EVI = \left[\text{Revenue}_i - \text{Revenue}_c - (\text{AcqExp} + \text{AdmExp} + \text{AppExp}) \right] * \frac{T}{t},$$

где:

Revenue_i (доход_i) – доход, полученный с использованием информационного актива (информированная группа);

Revenue_c (доход_c) – доход, полученный без использования информационного актива (контрольная группа);

T – средняя ожидаемая продолжительность жизни записи;

t – период времени, в течение которого проводился эксперимент или испытание EVI.

Реализация

Модель EVI как финансовый вариант модели PVI, описанной выше, требует проведения контролируемого эксперимента в течение определенного периода времени. Однако в этой модели доход – единственный ключевой показатель эффективности, значение представлено в денежном выражении, а не в виде коэффициента и учитывается продолжительность жизни информационного актива. Сначала измеряется разница между величинами дохода, генерируемыми с использованием информационного

актива и без использования. Затем вычитаются затраты на поддержку жизненного цикла информации (здесь можно использовать рекомендации модели CVI). Наконец, полученная величина умножается на отношение продолжительности жизни информационного актива (Т) к продолжительности испытания (t). При определении показателя EVI важно, чтобы все другие аспекты процесса получения дохода оставались во время испытания неизменными.

Преимущества и проблемы

Метод EVI – это эмпирическое исследование вклада информационного ресурса в соотношение дохода и расхода. Он не предполагает проведения функционального анализа, за исключением отдельных случаев определения расходов (связанных с информацией), когда данные дублируются и/или применяются несколькими способами. В то же время при использовании EVI требуется проведение эксперимента. Кроме того, необходима способность оценивать затраты на цепочку поставок информации. Следует заметить, что многие традиционные бизнес-лидеры все еще испытывают дискомфорт от современной концепции экспериментов с процессами, приносящими доход. В целом EVI – запаздывающий показатель, хотя результаты могут быть использованы для определения приоритетов ИТ- и бизнес-инициатив.

Описанные в этом разделе модели ни в коем случае не являются единственными методами оценки информационных активов. В различных источниках можно найти много других заслуживающих внимания подходов^{199, 200, 201, 202, 203}.

5.5. Три уровня ценности информационных активов

Затратный подход (CVI) позволяет определить лишь базовую, номинальную и консервативную ценность информации, независимо от того, как она используется или может быть использована. CVI соответствует способу, который бухгалтеры предпочитают применять для первоначальной оценки большинства нематериальных активов. Следовательно, цель организации должна состоять в том, чтобы найти способы получения для информационного актива рыночной ценности (MVI) или экономической ценности (EVI), которые бы превышали его CVI.

С конца 1980-х годов в международной практике бухгалтерского учета ценность актива можно регистрировать как отражающую его вероятные будущие экономические выгоды. Это означает, что неиспользуемый в настоящее время актив может иметь формальную измеримую ценность. Это также означает, что у любого актива есть три уровня ценности (рис. 5.9):

- реализованная ценность, основанная на экономических выгодах, которые информационные активы приносят сейчас;
- вероятная ценность, основанная на предполагаемом использовании;
- потенциальная ценность (больше теоретическая), если бы актив использовался оптимальным

¹⁹⁹ Black S. H., Marchand D. Assessing the Value of Information in Organisations: A Challenge for the 1980's // The Information Society Journal, 1982, 1 (3): 191–225.

²⁰⁰ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

²⁰¹ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

²⁰² Moody D., Walsh P. Measuring the value of information: An asset valuation approach // Proceedings of the 7th European Conference on Information Systems, Copenhagen, Denmark, 1999.

²⁰³ Wilson R. M. S., Stenson J. Valuation of information assets on the balance sheet: The recognition and approaches to the valuation of intangible assets // Business Information Review, 2008, 25 (3): 167–182. DOI: 10.1177/0266382108095039 – URL: <https://journals.sagepub.com/doi/abs/10.1177/0266382108095039?journalCode=bira>.

образом.

Для информационных активов подразумевается, что даже «темные данные» имеют как вероятную, так и потенциальную ценность. Поэтому утверждение, что информация имеет ценность только тогда, когда потребляется, не совсем верно.

Эти три уровня ценности также дают организациям возможность применить предлагаемые Gartner модели оценки информационных активов для выявления ценностных разрывов: в отношении эффективности и в отношении видения²⁰⁴.

1. Ценностный разрыв в отношении эффективности – это разница между вероятной и реализованной ценностью информационного актива.

Этот разрыв показывает, каким образом информация вероятнее всего принесет пользу. Например, данные технического обслуживания в настоящее время могут использоваться в организации для выявления неисправного оборудования и сокращения времени его простоя, но системы, которые позволят прогнозировать сбои, только создаются. Как модели MVI, так и модели EVI могут использоваться для фактической и прогнозной оценки разрыва в отношении эффективности. После измерения этого разрыва цель организации должна состоять в том, чтобы ускорить его ликвидацию и повысить вероятность использования информации должным образом.

2. Ценностный разрыв в отношении видения информации – разница между потенциальной и вероятной оценками информации.



Рис. 5.9.

Три уровня ценности информационных активов*

* Laney D. B. *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. *Инфономика: информация как актив: монетизация, оценка, управление*. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Этот разрыв позволяет оценить, в какой степени организация обладает видением того, как могут использоваться ее данные. Например, после внедрения систем технического обслуживания на основе прогнозов, существуют определенные дополнительные возможности для лицензирования предоставляемых этими системами данных поставщикам комплектующих в обмен на выгодные цены, а также ряд других потенциальных способов использования таких данных. Этот разрыв обычно намного больше, чем разрыв в отношении эффективности, но он также и более гипотетический. Каким же образом организация может определить все возможные способы использования информационного ресурса? Один из ответов состоит в том, чтобы определить фактическую и потенциальную значимость информационных активов для бизнеса (BVI), а затем провести мероприятия по инновациям в области данных, в которых основное внимание следует уделить тем информационным активам, у которых наибольшие разрывы. Теперь на основе полученных сведений о ранее не установленных потенциальных применениях информационного актива можно планировать их реализацию, тем самым увеличивая вероятную ценность

²⁰⁴ Laney D. B. *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. *Инфономика: информация как актив: монетизация, оценка, управление*. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

информационного актива и, в свою очередь, рыночную стоимость самой организации.

Литература к главе 5

- Oppenheim C., Stenson J., Wilson R. M. S. Studies on information as an Asset I: Definitions // Journal of Information Science, 2003, 29 (3): 159–166. DOI: 10.1177/0165551503029300.
- Oppenheim C., Stenson J., Wilson R. M. S. Studies on information as an Asset III: Views of information professionals // Journal of Information Science, 2004, 30 (2), 181–190. DOI: 10.1177/0165551504042809 – URL: https://repository.lboro.ac.uk/articles/journal_contribution/Studies_on_Information_as_an_Asset_III_Views_of_Information_Professionals/9413186/1.

Глава 6. Данные как актив: барьеры и ошибки на пути извлечения ценности

6.1. Барьеры на пути развертывания информационных активов и Лидерский манифест о данных

В предыдущей главе мы говорили о том, что переход к управлению данными как активом позволяет в наибольшей степени раскрыть их потенциальную ценность для организации. Что может помешать такому переходу?

В 2012 году австралийские специалисты Нина Эванс (Nina Evans) и Джеймс Прайс (James Price) опубликовали статью «Барьеры на пути эффективного развертывания информационных ресурсов: Точка зрения исполнительного руководства»²⁰⁵. В ней они представили результаты исследования, проведенного с привлечением исполнительных руководителей крупных австралийских и южноафриканских компаний. Эта часто цитируемая работа представляет собой аналитический отчет, систематизирующий сведения об основных барьерах, с которыми сталкиваются организации на пути развертывания своих информационных активов.

Как замечают авторы исследования, многие программы руководства в области данных²⁰⁶ создаются исходя исключительно из требований по обеспечению нормативно-правового соответствия (compliance), а не из соображений реализации потенциальных возможностей по извлечению ценности из данных как актива. Недопонимание руководителями важности управления данными как активом приводит к недостаточной приверженности этой деятельности внутри организации, включая деятельность по управлению качеством данных..

Результаты Эванс и Прайса были взяты за основу и существенным образом дополнены авторами Лидерского манифеста о данных (далее мы рассмотрим его более подробно)²⁰⁷.

В качестве корневых факторов, препятствующих эффективному управлению данными как активом, были выделены следующие:

- непонимание значения управления данными руководством и сотрудниками;
- недостаточная управляемость бизнеса;
- недостатки в руководстве и оперативном управлении;
- трудности с обоснованием необходимости совершенствования управления данными;
- использование неподходящих или неэффективных инструментов.

²⁰⁵ Evans N., Price J. Barriers to the Effective Deployment of Information Assets: An Executive Management Perspective // Interdisciplinary Journal of Information, 2012, 7: 177–199. DOI: 10.28945/1721. – URL: https://www.researchgate.net/publication/289400090_Barriers_to_the_Effective_Deployment_of_Information_Assets_An_Executive_Management_Perspective.

²⁰⁶ Тему руководства данными (data governance) мы подробно обсудим в главе 10.

²⁰⁷ McGilvray D., Price J., Redman T. Barriers that slow, hinder, prevent companies from managing their information as a business asset, 2016. – URL: <https://dataleaders.org/tools/root-cause-analysis/>.

Факторы представлены на рисунке 6.1.

В 2017 году группа ведущих мировых экспертов в области управления данными во главе с Джоном Лэдли опубликовала Лидерский манифест о данных²⁰⁸. В нем, в частности, говорится, что «лучшие возможности для органичного роста организации заложены в данных». Хотя в большинстве организаций признают за данными статус актива, компании все еще далеки от того, чтобы называть себя «управляемыми на основе данных». Более того, большинство из них даже не представляют, какими данными владеют и какие именно данные имеют решающее значение для их бизнеса. Организации продолжают не видеть разницы между данными и информационными технологиями и плохо управляют как тем, так и другим. Такое положение дел усугубляет проблемы управления данными и подчеркивает критически важный фактор потенциального успеха организации: лидерство и приверженность руководства, умноженные на вовлечение всех без исключения сотрудников на всех уровнях организации.

Авторы манифеста призывают все заинтересованные стороны принять участие в устранении имеющихся барьеров в налаживании в своих организациях практики управления информационными активами.

6.2. Дата-центричное мышление

В предыдущем разделе мы отметили основные барьеры, препятствующие эффективному управлению данными как активом. Теперь остановимся на основных ошибках, которые могут помешать организациям преодолеть эти барьеры и внедрить устойчивую практику управления данными. Какие же ошибки могут быть допущены?

Известный бизнес-гуру Питер Айкен (Peter Aiken)²⁰⁹ в своих публикациях приводит семь «смертных грехов» в области работы с данными (The Seven Deadly Data Sins)^{210, 211}.

²⁰⁸ Ladley J., McGilvray D., O'Neal K., Price J., Redman T. The Leader's Data Manifesto. dataleaders.org, 2019. – URL: <https://dataleaders.org/>.

²⁰⁹ Питер Айкен – один из наиболее авторитетных мировых экспертов по управлению данными, бывший президент Международной ассоциации управления данными (DAMA), основатель и директор Data Blueprint – консалтинговой фирмы с хорошей репутацией, постоянный докладчик на отраслевых форумах, преподаватель, автор нескольких книг.

²¹⁰ Aiken P. The Seven Deadly Data Sins – Emerging from Management Purgatory. Dataversity, 2017. – URL: <https://www.dataversity.net/data-ed-slides-seven-deadly-data-sins-emerging-management-purgatory/>.

²¹¹ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

ЛИДЕРСКИЙ МАНИФЕСТ О ДАННЫХ

ВОЗМОЖНОСТИ ДЛЯ РОСТА ВАШЕЙ ОРГАНИЗАЦИИ КРОЮТСЯ В ВАШИХ ДАННЫХ

ДАННЫЕ ХРАНЯТ В СЕБЕ ОГРОМНЫЙ ПОТЕНЦИАЛ ДЛЯ СОЗДАНИЯ КОНКУРЕНТНОГО ПРЕИМУЩЕСТВА, БИЗНЕС-ЦЕННОСТЕЙ И РАБОЧИХ МЕСТ; УЛУЧШЕНИЯ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ; ПОВЫШЕНИЯ ВАШЕЙ БЕЗОПАСНОСТИ; И В ЦЕЛОМ УЛУЧШЕНИЯ ВАШЕГО СОСТОЯНИЯ.

ОРГАНИЗАЦИЯМ ЕЩЕ ДАЛЕКО ДО СОСТОЯНИЯ, КОГДА ВСЕ РЕШЕНИЯ БУДУТ ОСНОВАНЫ НА ДАННЫХ.

БОЛЬШИНСТВО КОМПАНИЙ:



- 1) НЕ ЗНАЮТ, КАКИЕ ДАННЫЕ ДОСТУПНЫ ВНУТРИ ОРГАНИЗАЦИИ И КАКИЕ ИЗ НИХ НАИБОЛЕЕ ЦЕННЫЕ
- 2) ПУТАЮТ ПОНЯТИЕ «ДАННЫЕ» С ПОНЯТИЕМ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ», ЧТО ВЕДЕТ К НЕПРАВИЛЬНОМУ ТРАКТОВАНИЮ ОБОИХ ПОНЯТИЙ
- 3) НЕ ИМЕЮТ СТРАТЕГИИ ПО ИСПОЛЬЗОВАНИЮ ДАННЫХ И ПОНИМАНИЯ ЦЕННОСТИ ДАННЫХ ДЛЯ ОРГАНИЗАЦИИ И БИЗНЕСА
- 4) НЕ ИМЕЮТ РЕСУРСОВ И НЕДООЦЕНИВАЮТ УСИЛИЯ, НЕОБХОДИМЫЕ ДЛЯ УПРАВЛЕНИЯ ДАННЫМИ

БОЛЬШИНСТВО КОМПАНИЙ УСПЕШНЫ С НЕБОЛЬШИМИ ПРОЕКТАМИ ПО УПРАВЛЕНИЮ ДАННЫМИ, ОБЕСПЕЧЕНИЮ КАЧЕСТВА И БЕЗОПАСНОСТИ ДАННЫХ, АНАЛИТИКЕ И ДРУГИМ ИНИЦИАТИВАМ, СВЯЗАННЫМ С ДАННЫМИ. ДО СИХ ПОР МЫ НЕ ВСТРЕЧАЛИ ПРИМЕРОВ **ФУНДАМЕНТАЛЬНЫХ, ПРОДОЛЖИТЕЛЬНЫХ ИЗМЕНЕНИЙ В КОМПАНИИ БЕЗ ОБЯЗАТЕЛЬНОГО УЧАСТИЯ РУКОВОДСТВА И УЧАСТИЯ ВСЕХ И КАЖДОГО НА ВСЕХ УРОВНЯХ ОРГАНИЗАЦИИ.**

...И МЫ ПОЛНОСТЬЮ ПОНИМАЕМ, КАК СЛОЖНО РАСКРЫТЬ ПОТЕНЦИАЛ ДАННЫХ ДЛЯ ВСЕЙ ОРГАНИЗАЦИИ

— ПОЭТОМУ МЫ ПРИЗЫВАЕМ ВСЕХ ИЗМЕНИТЬ СВОЕ ОТНОШЕНИЕ К ДАННЫМ И ЗНАНИЯМ О НИХ —

ЧЛЕНЫ ПРАВЛЕНИЯ, РУКОВОДИТЕЛИ И ЛИДЕРЫ ОРГАНИЗАЦИЙ: БРОСЬТЕ ВЫЗОВ СВОИМ УСТОЯВИМЫМ ПОНЯТИЯМ О ДАННЫХ.

Воспринимайте данные не как элементы, скрытые в недрах ИТ, а как источник неограниченных возможностей. Поймите, что данные полезны не только для специалистов по работе с данными, таких как аналитики и разработчики, но и для всех в вашей компании, чтобы действительно отличаться от конкурентов.

Рассматривайте данные как ваше личное наследие. Подумайте, что бы значило для вас размещение активов данных на балансе вашей организации.

Лидерам необходимо выработать видение для данных. В качестве первого шага сфокусируйтесь на следующих областях:

1. Лучше заботьтесь о данных, уделяя особое внимание качеству наиболее важных данных.
2. Испытайте множество способов использования ваших данных и получения ценности и конкурентного преимущества.
3. Усовершенствуйте систему управления, более подходящую для управления данными.

ВСЕ, КТО НУЖДАЮТСЯ В ДАННЫХ ДЛЯ ВЫПОЛНЕНИЯ СВОЕЙ РАБОТЫ, СТАНЬТЕ ПРОВОКАТОРАМИ ДАННЫХ.

Возможностей предостаточно, поэтому выберите интересующую вас область. Это может быть улучшение качества данных, обнаружение более глубокого аналитического метода, разработка новой метрики, выдвижение идеи, которая количественно оценивает ценность данных в рублях, или использование данных для выстраивания отношений.

ПРОФЕССИОНАЛЫ ПО РАБОТЕ С ДАННЫМИ, БУДЬТЕ ПРОАКТИВНЫМИ.

Общайтесь со своими коллегами по бизнесу, продавайте концепции, становитесь наставником данных и помогайте людям создавать собственные истории успеха.

ПРИСТУПАЙТЕ К РАБОТЕ.

Это захватывающие и рискованные времена. Захватывающие, потому что данные предоставляют возможности для создания конкурентного преимущества, улучшения существующих продуктов и услуг (и создания новых), лучшего понимания клиентов и снижения затрат. И рискованные, потому что исправление того, что сломано или работает неправильно, будет трудным, и те, кто ждет слишком долго и ничего не делает, могут оказаться в крайне неблагоприятном положении.

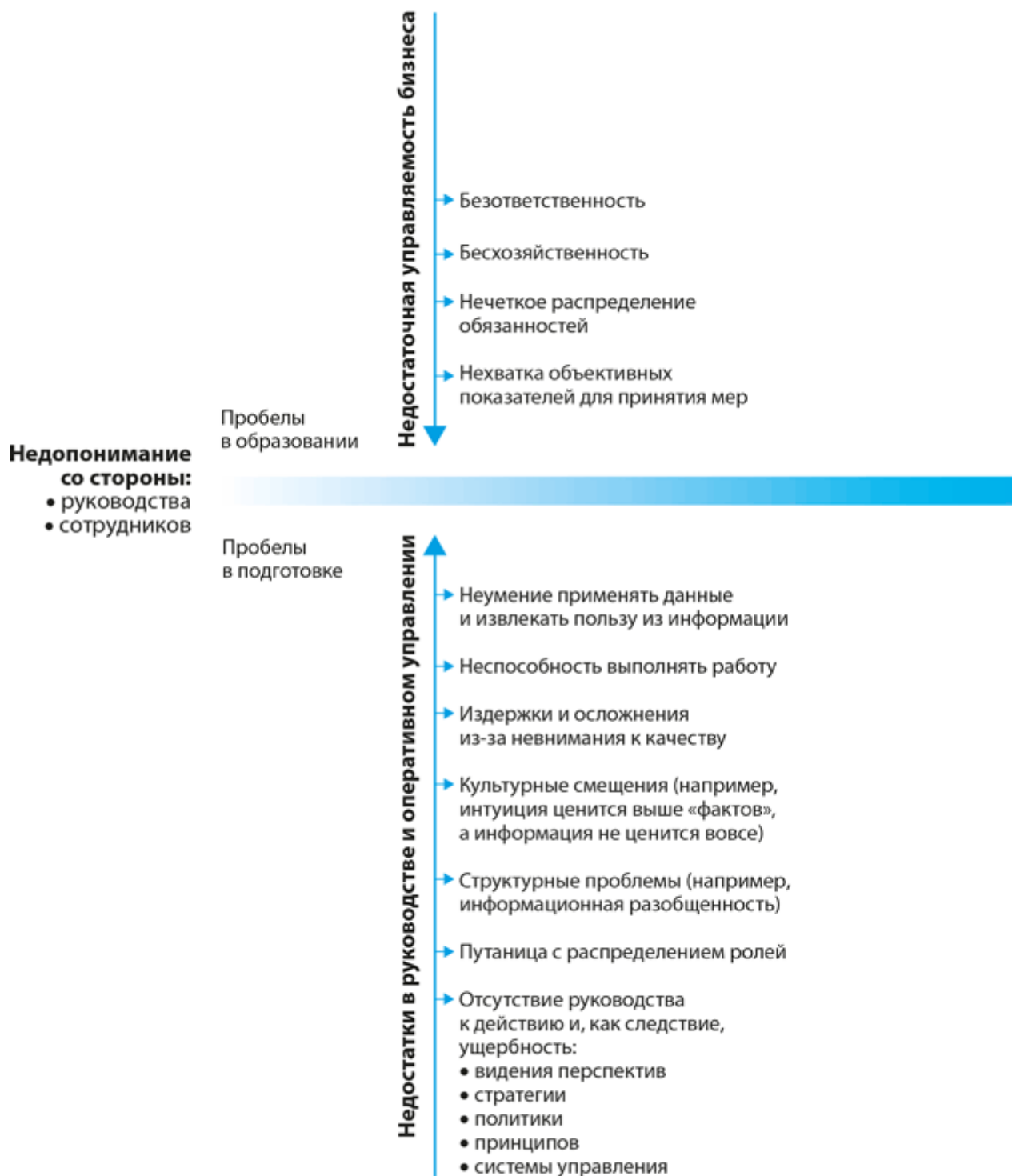
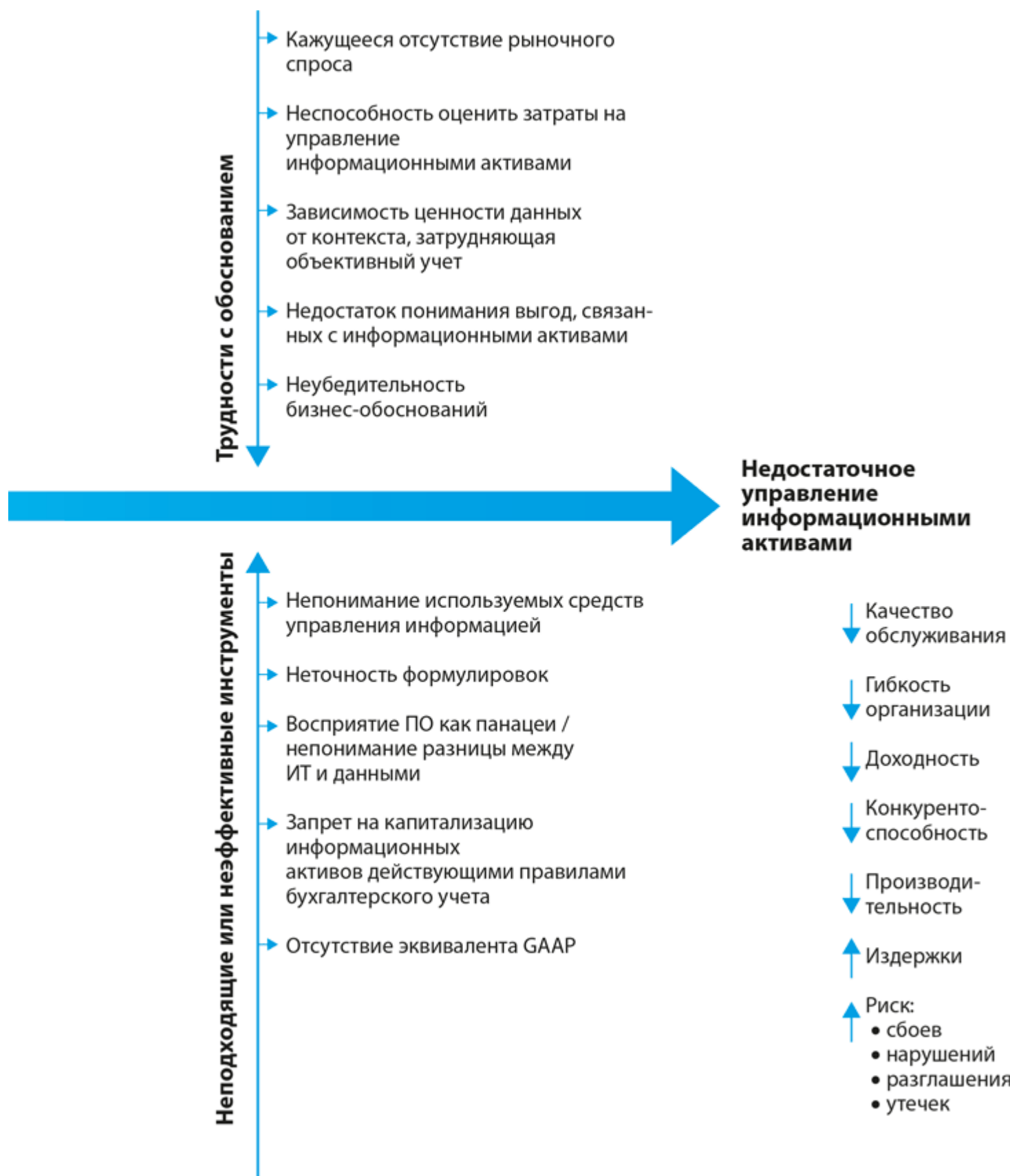


Рис. 6.1.

Факторы, мешающие организациям эффективно управлять данными как активом*



* McGilvray D., Price J., Redman T. Barriers that slow, hinder, prevent companies from managing their information as a business asset, 2016. – URL: <https://dataleaders.org/tools/root-cause-analysis/>.

1. Непонимание основ дата-центричного мышления.
2. Отсутствие квалифицированного руководства и лидерства в области данных.
3. Неспособность внедрить программный подход к организации совместного использования данных.

4. Отсутствие согласованности программы в области данных с ИТ-проектами.
5. Неспособность адекватно управлять ожиданиями.
6. Отсутствие последовательности в реализации стратегии в области данных.
7. Отсутствие внимания к проблемам, связанным с культурой и управлением изменениями.

Что такое дата-центричное мышление (data-centric thinking) и почему непонимание его основ поставлено на первое место в ряду перечисленных «смертных грехов»?²¹².

В книге «Стратегия обработки данных и корпоративное руководство в области данных: Обеспечение синхронизации бизнеса и ИТ в эпоху после эпохи больших данных» Айкен и его коллега Тодд Харбор (Todd Harbour) обращают внимание на возвышенные слова из действующего в Европейском союзе Общего регламента по защите данных (General Data Protection Regulation, GDPR): «Обработка персональных данных должна быть направлена на служение человечеству»²¹³.

Причина подобного решительного заявления проста. Становится все более очевидным, что ни один аспект нашей деятельности не застрахован от негативных последствий плохо организованной работы с информационными активами. Последствия могут быть многочисленными и далеко идущими. Достаточно подумать о затратах многих организаций на постоянную проверку своих информационных систем в поисках цифрового мусора и сомнительных материалов. Здесь можно провести аналогию с финансовыми вложениями, которые организации должны делать, чтобы противодействовать постоянно растущему потоку ненужных и мешающих отходов.

В связи с возрастающим значением эффективной работы с данными в последнее время все чаще говорится о дата-центричном мышлении, или мышлении, ориентированном на данные (data-centric thinking). Прежде чем попытаться определить это понятие, рассмотрим часто наблюдаемые последствия пренебрежительного отношения организаций к основополагающей роли данных и их неготовности перейти к такому образу мышления²¹⁴.

- Недостаточная грамотность в области данных (data literacy) на всех уровнях организации приводит к неполному пониманию сотрудниками ценности совместно используемых информационных активов. Это заставляет организации сосредоточиться на такой более простой с точки зрения концептуального восприятия деятельности, как разработка и внедрение программного обеспечения и базовых информационно-технологических решений.

- Недостаточное понимание роли данных приводит к тому, что организации игнорируют потребность в целостных полномасштабных программах в области данных и вместо этого пытаются управлять совместно используемыми информационными активами на уровне отдельных проектов.

- Отсутствие программ в области данных приводит к увеличению расходов на ИТ. Организации тратят ресурсы на такие виды деятельности, как интеграция и очистка данных или управление гораздо большим количеством данных, чем это необходимо для решения их стратегических задач.

- Отсутствие возможности подготовиться к будущим изменениям путем внедрения гибкой и адаптируемой архитектуры корпоративных данных также приводит к излишнему расходу ресурсов.

- Слабое представление о способности информационных активов эффективно и действенно поддерживать стратегию организации приводит к снижению эффективности ее деятельности.

- Большие объемы неуправляемых данных увеличивают сложность процессов внутри организации.

- Увеличение количества времени, усилий и рисков, связанных с ИТ-проектами, угрожает снижением прибыли.

- Неспособность обеспечить гибкость и адаптивность архитектуры данных до начала функционирования процессов основной деятельности организации требует дополнительного времени и финансирования для устранения связанных с этим негативных последствий.

- Отсутствие возможности создавать повторно используемые решения, ориентированные на данные,

212 Все «смертные грехи» мы обсудим в следующем разделе.

213 Там же.

214 Там же.

требует дублирования усилий, снижает качество и надежность информации и стоит денег.

- Увеличение времени, затрачиваемого на достижение понимания данных, и соответствующее сокращение времени и затрат на анализ.
- Недостаточное понимание информационных активов препятствует любой возможности рассматривать (а тем более реализовывать) элементы стратегии организации, ориентированные на данные.
- Снижение уверенности в правильности принимаемых решений – неблагоприятный результат непонимания информационных активов организации.
- Наконец, – и, возможно, это самое неприятное – возрастание объемов излишних, устаревших и тривиальных данных (data ROT)²¹⁵ приводит к тому, что ситуация по всем перечисленным аспектам только ухудшается.

Такое явление, как data ROT, стоит рассмотреть более подробно – слишком много данных низкого качества являются избыточными и не соответствуют определению информационных активов.

Данные, как и многие другие ресурсы, имеют жизненный цикл – это означает, что в какой-то момент они могут утратить свою ценность и более не соответствовать первоначальному назначению. В наше время, когда стоимость технологий находится в относительно доступных пределах, технически несложно генерировать отдельные копии данных для широкого спектра индивидуальных целей. При этом объем данных растет с такой поразительной скоростью и благодаря такому разнообразию источников, которые руководители организаций даже не могут себе представить.

Сотрудники часто создают копию данных, используют ее для выполнения частной задачи, а затем их внимание переключается на другую задачу. Получив новое задание, люди быстро забывают о сгенерированных ими данных и никогда их больше не используют.

Это влечет за собой появление того, что называется «темными данными» (dark data)²¹⁶. «Темные данные» можно рассматривать как особую разновидность data ROT.

Хотя термин «дата-центричное мышление» используется достаточно часто, у него до сих пор отсутствует какое-то устойчивое общепринятое определение. Айкен и Харбор решили заняться выработкой коллективного понимания основ дата-центричного мышления и определением конкретных шагов, необходимых для достижения лучших результатов при работе организаций с данными. В 2017 году они опубликовали так называемую Доктрину в области данных, которая в дальнейшем была откорректирована на основе откликов, поступивших от сотен представителей заинтересованного профильного сообщества. Вторая версия доктрины была опубликована в 2021 году²¹⁷.

Доктрина в области данных – не единственный известный концептуальный документ, направленный на пропаганду идей дата-центричности. На сайте доктрины представлены ссылки на похожие по тематике и направленности тексты:

- Манифест дата-центричности (The Data-Centric Manifesto), разработанный командой американской

²¹⁵ В англоязычных источниках по отношению к излишним, устаревшим и тривиальным данным часто используется акроним ROT (Redundant, Obsolete, Trivial) и термин data ROT*. На русский язык его можно перевести как «информационная гниль» (от англ. rot – «гниль», «труха»).

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

²¹⁶ Gartner определяет «темные данные» как информационные активы, которые организации собирают, обрабатывают и хранят, но не используют*.

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

²¹⁷ Aiken P., Harbour T. The Data Doctrine™ V2. Objective measures for improving data outcomes. Data Blueprint Inc., 2021. – URL: <https://anythingawesome.com/TheDataDoctrineV2.html>.

консалтинговой компании Semantic Arts^{218219, 220};

- Лидерский манифест о данных²²¹.

Доктрина в области данных построена по аналогии со знаменитым манифестом гибкой разработки программного обеспечения (Agile-манифест), который был опубликован в 2001 году²²²²²³.

Agile-манифест разработки программного обеспечения

Мы постоянно открываем для себя более совершенные методы разработки программного обеспечения, занимаясь разработкой непосредственно и помогая в этом другим. Благодаря проделанной работе мы смогли осознать, что:

- люди и взаимодействие важнее процессов и инструментов;
- работающий продукт важнее исчерпывающей документации;
- сотрудничество с заказчиком важнее согласования условий контракта;
- готовность к изменениям важнее следования первоначальному плану.

Не отрицая важности того, что справа, мы все-таки больше ценим то, что слева.

6.3. Доктрина в области данных (версия 2)

Целевые меры для повышения результативности работы с данными

Мы постоянно открываем для себя более совершенные методы разработки информационных систем, занимаясь разработкой непосредственно и помогая в этом другим. Благодаря проделанной работе мы смогли осознать, что:

- **программы в области данных** играют ведущую роль по отношению к программам в области ИТ;
- **обоснованное инвестирование в информационное обеспечение** важнее деятельности по приобретению технологий;
- **стабильные по структуре и совместно используемые данные организации** важнее совершенствования технологических компонентов;
- **повторное использование данных** важнее новых источников данных.

Не отрицая важности того, что справа, мы все-таки больше ценим то, что слева.

6.4. Доктрина в области данных: базовые ценности дата-центричного мышления

Доктрина в области данных включает четыре базовых утверждения относительно ценностных приоритетов, на которых должна основываться разработка информационных систем в организации, ориентированной на дата-центричное мышление. Рассмотрим их более подробно.

218 Манифест дата-центричности обсуждается в главе 7.

219 McComb D. The Data-Centric Revolution. The Data Administration Newsletter (TDAN.COM), 2015. — URL: <https://tdan.com/the-data-centric-revolution/18780>.

220 Semantic Arts. The Data-Centric Manifesto. Semantic Arts Inc., 2015. — URL: <http://datacentricmanifesto.org/>.

221 Ladley J., McGilvray D., O'Neal K., Price J., Redman T. The Leader's Data Manifesto. dataleaders.org, 2019. — URL: <https://dataleaders.org/>.

222 Agile-манифест создан группой из 17 независимых практиков в области разработки программного обеспечения, назвавшей себя Agile Alliance. Он включает 4 ценности и 12 принципов. Текст манифеста доступен на более чем 50 языках (в т. ч. на русском).

223 Agile Alliance. Manifesto for Agile Software Development, 2001. — URL: <https://agilemanifesto.org/iso/ru/manifesto.html>.

Ценность № 1. Программы в области данных играют ведущую роль по отношению к программам в области ИТ

Здесь имеется в виду определяющая позиция программ проектов и мероприятий в области архитектуры, моделирования, проектирования, описания и подготовки данных организации по отношению к программе ИТ-проектов (в данном случае под ИТ-проектами подразумеваются проекты по разработке и внедрению программных приложений и базовых информационно-технологических решений).

Чтобы извлечь максимальную ценность из своих данных, организации должны подойти к ним как к активу и организовать этот актив (наряду с другими своими активами) таким образом, чтобы он способствовал достижению их стратегических целей. С ростом стратегической ценности данных возрастают и возможности по их повторному использованию во многих программных приложениях. Для реализации таких возможностей организации должны сознательно направлять усилия по обеспечению использования доступных стратегически важных данных (работа с которыми была налажена в результате уже выполненных проектов по разработке приложений) теми приложениями, которые создаются в рамках новых ИТ-проектов. Связанную с этим деятельность необходимо осуществлять целенаправленно и непрерывно, распространяя на все проекты по разработке приложений. Таким образом, прилагаемые усилия по обеспечению повторного использования данных – это не отдельный проект (имеющий завершение), а программа – деятельность, которая инициируется и продолжается до тех пор, пока организация не решит, что в ее осуществлении больше нет необходимости. Программа в области данных должна быть организационно оформлена отдельно от программ ИТ-проектов. Ее задача – реализация единой корпоративной стратегии работы с данными, которая обеспечит максимальную информационную поддержку стратегии бизнеса.

На рисунке 6.2 показано, каким образом увеличивается использование данных организации в результате увязки отдельных ИТ-проектов с программой в области данных. В рамках программы вырабатываются требования по архитектуре данных, реализуемые в ходе индивидуальных ИТ-проектов. По завершении каждого проекта область данных, доступных для совместного использования многими приложениями, расширяется за счет реализованного в проекте нового организованного сегмента.

С каждым ИТ-проектом нарабатывается опыт реализации стратегии в области данных, выявляются новые потребности в данных и совершенствуется координация программы в области данных и программ ИТ. Поэтому объемы используемых и предоставляемых очередным приложением данных возрастают (на рисунке это отражено как увеличение толщины соответствующих стрелок).

К сожалению, в настоящее время в большинстве организаций распространена практика разработки информационных систем с архитектурой, ориентированной на приложения. Архитектура данных в таких системах создается в расчете на использование только в рамках одного приложения и непригодна для использования другими системами^{224, 225}. Избавиться от недостатков ориентации на приложения и организовать постепенный переход к дата-центричной архитектуре можно, обеспечив последовательное влияние на ИТ-проекты с помощью программы в области данных.

²²⁴ McComb D. The Data-Centric Revolution: Data-Centric vs. Data-Driven. The Data Administration Newsletter (TDAN.COM), 2016. – URL: <https://tdan.com/the-data-centric-revolution-data-centric-vs-data-driven/20288>.

²²⁵ McComb D. The Data-Centric Revolution. The Data Administration Newsletter (TDAN.COM), 2015. – URL: <https://tdan.com/the-data-centric-revolution/18780>.

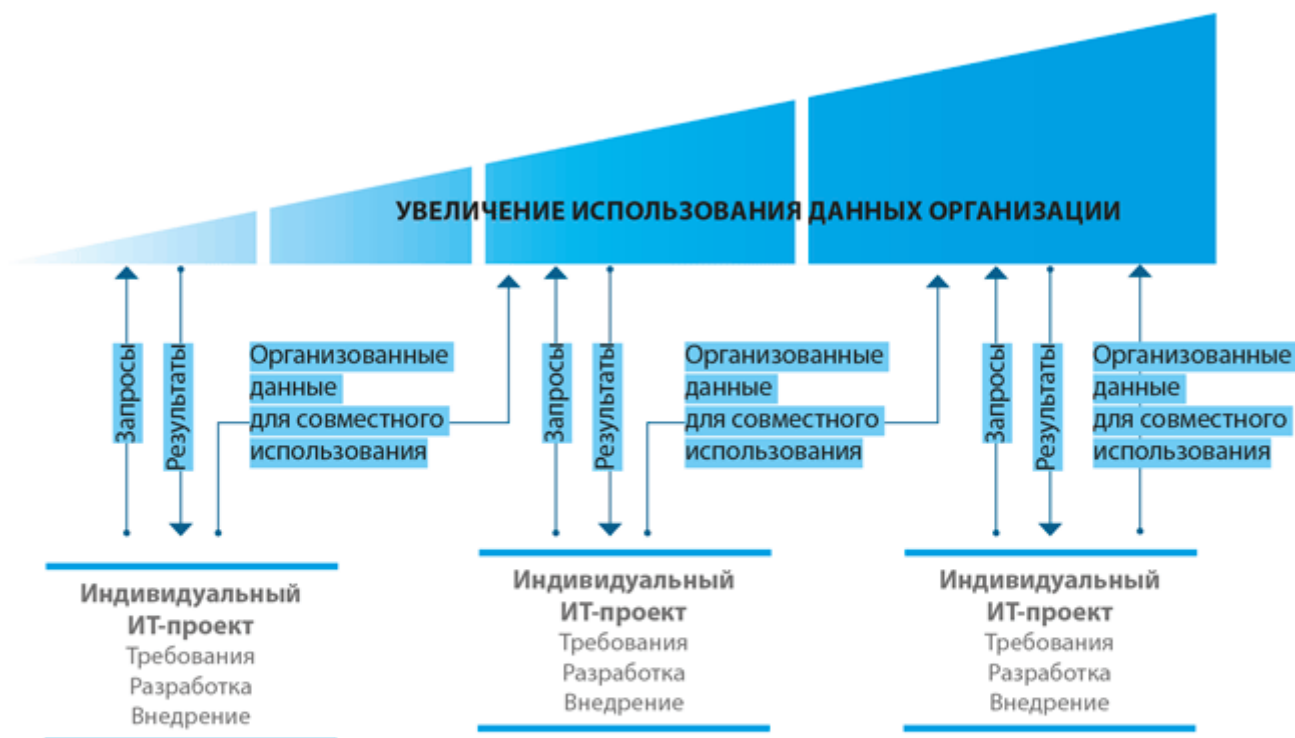


Рис. 6.2.

Увеличение использования данных организации в результате увязки программы в области ИТ с программой в области данных*

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Ценность № 2. Обоснованное инвестирование в информационное обеспечение важнее деятельности по приобретению технологий

Выше мы уже перечисляли последствия игнорирования организациями основополагающей роли данных и их неготовности перейти к дата-центричному образу мышления.

Тут можно только повторить, что отсутствие проектных программ в области данных, обоснованно увязанных с бизнес-стратегией организации, приводит к увеличению расходов на ИТ. Без четкого понимания потребностей в информационном обеспечении деятельности организации приобретаемые технологии могут привести даже к снижению ее эффективности. Это может произойти за счет добавления дополнительных операций по интеграции и очистке данных или организации управления гораздо большим количеством данных, чем это необходимо для решения стратегических задач.

Например, организация с целью решения проблем своевременного обеспечения сотрудников качественными, совместно используемыми важными данными может реализовать проект по внедрению дорогостоящей MDM-платформы²²⁶. Однако через некоторое время после завершения проекта выясняется, что ситуация не улучшилась. Хотя ИТ-команда и выполнила проект в соответствии с техническим заданием, но организация не получила ожидаемую пользу от внедренной технологии. Это произошло, потому что не был должным образом проведен целый ряд необходимых предварительных мероприятий: определение общеорганизационных требований и разработка согласованных решений в части системы классификации и кодирования, правил описания учитываемых позиций, правил обеспечения качества; определение надежных, легитимных источников данных; подготовка данных; решение организационных вопросов и т. п.

226 MDM (Master Data Management) – управление основными данными.

Как замечают Айкен и Харбор, когда организации покупают высокотехнологичные системы без соответствующей подготовки, они не понимают, что это все равно что дать начинающему водителю ключи от автомобиля Tesla, полагая, что он знает, как управлять этим сложным инновационным транспортным средством²²⁷.

Ценность № 3. Стабильные по структуре, совместно используемые данные организации важнее совершенствования технологических компонентов

Организации в рамках своих ИТ-проектов редко формируют структуры данных в расчете на их повторное использование, т. е. в виде хорошо продуманных и документированных моделей. Такая ситуация сложилась в силу распространения подхода к разработке информационных систем, ориентированного на приложения^{228, 229}. Отчасти оказывает влияние и слишком буквальное следование принципам Agile-манифеста²³⁰. Подобная практика (особенно если заранее предполагается, что данные будут использоваться несколькими приложениями) создает много проблем: удорожание и увеличение сроков разработки, меньшую отдачу от проектов, увеличение рисков. Она не учитывает, что изменения в структуре совместно используемых данных могут повлиять на каждое получающее к ним доступ программное приложение.

Если организации планируют использовать данные повторно, они должны руководствоваться принципами проектирования, которые способствуют многократному применению. Структуры данных должны быть стабилизированы до того, как начнут создаваться или развиваться приложения, обращающиеся к ним.

Ценность № 4. Повторное использование данных важнее получения новых источников данных

Организациям необходимо разработать процессы, обеспечивающие сбор и поддержание в актуальном состоянии требований к данным, сведений об имеющихся источниках данных и процедурах работы с ними, которые должны учитываться при разработке приложений.

Планомерность такой деятельности (включая подробное документирование) способствует выработке общего понимания фундаментальных концепций работы с данными и более точной оценке требуемых ресурсов и затрат. Создание программных приложений осуществляется согласованно, на основе общих стабильных структур данных, спецификаций и документации.

Однако далеко не все организации работают подобным образом. Отсутствие хорошо описанных стабильных структур и источников данных приводит к тому, что они определяются исходя из потребностей конкретного проекта. Это существенно осложняет интеграцию систем и организацию обмена данными, когда возникает такая необходимость.

6.5. Семь «смертных грехов» в области работы с данными

В начале главы мы перечислили семь «смертных грехов» в области работы с данными (рис. 6.3). Остановимся коротко на каждом из них^{231, 232}.

²²⁷ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

²²⁸ McComb D. The Data-Centric Revolution. The Data Administration Newsletter (TDAN.COM), 2015. – URL: <https://tdan.com/the-data-centric-revolution/18780>.

²²⁹ Semantic Arts. The Data-Centric Manifesto. Semantic Arts Inc., 2015. – URL: <http://datacentricmanifesto.org/>.

²³⁰ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

²³¹ Aiken P. The Seven Deadly Data Sins – Emerging from Management Purgatory. Dataversity, 2017. – URL: <https://www.dataversity.net/data-ed-slides-seven-deadly-data-sins-emerging-management-purgatory/>.

²³² Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big

1. Непонимание основ дата-центричного мышления

Идеи и мотивы, лежащие в основе дата-центричного мышления, требуют, чтобы соответствующие темы вошли в качестве базовых в программы высших и средних специальных учебных заведений. К сожалению, сейчас этому уделяется недостаточно внимания. От начала разработки программ обучения до массового выпуска подготовленных специалистов могут пройти долгие годы.

В настоящее время компаниям и учреждениям, чтобы заполнить пробел в знаниях и повысить зрелость специалистов по работе с данными, в основном приходится опираться на материалы наиболее авторитетных профессиональных организаций, таких как Международная ассоциация управления данными (Data Management Association International, DAMA)²³³, Институт моделирования зрелости способностей (Capability Maturity Model Institute, CMMI)²³⁴, Совет по управлению корпоративными данными (Enterprise Data Management Council, EDM Council)²³⁵ и ряд других²³⁶.

Data Era. Technics Publications, 2017.

²³³ <https://www.dama.org/cpages/home>.

²³⁴ <https://cmmiinstitute.com>.

²³⁵ <https://edmcouncil.org>.

²³⁶ В Российской Федерации в 2021 году начала работать официально зарегистрированная Ассоциация профессионалов управления данными (АПУД) (<http://dgpa-rus.org/>).

- 1 НЕПОНИМАНИЕ ОСНОВ ДАТА-ЦЕНТРИЧНОГО МЫШЛЕНИЯ
- 2 ОТСУТСТВИЕ КВАЛИФИЦИРОВАННОГО РУКОВОДСТВА И ЛИДЕРСТВА В ОБЛАСТИ ДАННЫХ
- 3 НЕСПОСОБНОСТЬ ВНЕДРИТЬ ПРОГРАММНЫЙ ПОДХОД К ОРГАНИЗАЦИИ СОВМЕСТНОГО ИСПОЛЬЗОВАНИЯ ДАННЫХ
- 4 ОТСУТСТВИЕ СОГЛАСОВАННОСТИ ПРОГРАММЫ В ОБЛАСТИ ДАННЫХ С ИТ-ПРОЕКТАМИ
- 5 НЕСПОСОБНОСТЬ АДЕКВАТНО УПРАВЛЯТЬ ОЖИДАНИЯМИ
- 6 ОТСУТСТВИЕ ПОСЛЕДОВАТЕЛЬНОСТИ В РЕАЛИЗАЦИИ СТРАТЕГИИ В ОБЛАСТИ ДАННЫХ
- 7 ОТСУТСТВИЕ ВНИМАНИЯ К ПРОБЛЕМАМ, СВЯЗАННЫМ С КУЛЬТУРОЙ И УПРАВЛЕНИЕМ ИЗМЕНЕНИЯМИ

Рис. 6.3.

Семь «смертных грехов» в области работы с данными

Как и другие предметы обучения, управление данными требует прочной, продуманной и всеобъемлющей образовательной основы. Нельзя построить какую-либо структуру на плохо спроектированном фундаменте, поэтому организациям необходимы надежные базовые методы управления данными.

В частности, компаниям и учреждениям необходимо стандартизировать способы получения, обработки, хранения и совместного использования информационных активов. В противном случае на всех уровнях управления будут продолжать изобретать свои собственные приемы и процедуры в ущерб организации в целом. Каждая организация и ее руководство обязаны обеспечить повышение грамотности корпоративной работы с информационными активами всех сотрудников, а не только команды по управлению данными. Это непростая задача, потому что в кругах руководства мало кто понимает основы управления данными.

Неосведомленность лидеров ставит организацию в сложное положение. Большинство руководителей организаций не осознают, что работа по управлению данными основана на особых дата-центричных подходах к архитектуре и разработке, и ее непросто сочетать с утвердившимся опытом создания программного обеспечения и другой проектной работы в области ИТ. Отсюда следуют два ошибочных представления, которые существенно влияют на увеличение затрат:

- управление данными рассматривается как часть отдельных ИТ-проектов;
- бизнес ожидает, что управление данными осуществляется правильным образом силами одного блока ИТ.

Современная практика состоит в том, что подготовленные специалисты по управлению данными должны на постоянной основе тесно сотрудничать с бизнесом. Основная цель такого сотрудничества – выявление и обобщение потребностей бизнеса с последующей их фиксацией в виде практических

требований к данным. Организациям необходимо инвестировать время, деньги и ресурсы для смещения фокуса с информационных технологий на сами данные.

2. Отсутствие квалифицированного руководства и лидерства в области данных

Как зарубежная, так и российская система высшего образования до сих пор рассматривают область данных как составную часть комплекса технических дисциплин по информационным технологиям. Отдельной представленной в программах вузов фундаментальной базы знаний, необходимой для управления данными как деятельностью, связанной с решением масштабных и сложных проблем на уровне организации, не существует. Вместо этого некоторые учебные заведения предлагают выборочные темы в рамках множества различных программ, в первую очередь по библиотечному делу, компьютерным наукам и управлению бизнесом.

Поскольку роль руководителя по работе с данными все еще не имеет четких академических представлений о предъявляемых к ней квалификационных требованиях, организации в значительной степени не осведомлены о бизнес-процессах, архитектуре и методах, необходимых для успешного повторного использования и оптимизации информационных активов.

По мнению ряда специалистов, в настоящее время помимо Свода знаний по управлению данными – DAMA-DMBOK²³⁷ и Модели зрелости управления данными Института CMMI–CMMI DMM²³⁸ практически не существует доступных полноценных материалов по управлению данными, которые не зависят от поставщика соответствующих решений²³⁹. Посодействовать в восполнении такого пробела в знаниях пока могут только профессиональные организации и консультанты.

Если рассматривать данные как корпоративный актив, можно провести некоторую аналогию между ролью руководителя по работе с данными и ролью финансового директора.

Финансовый директор прежде всего отвечает за управление финансовыми активами организации. Кроме того, он несет ответственность перед вышестоящим руководством за финансовое планирование и ведение учета, а также за финансовую отчетность. В некоторых секторах финансовый директор дополнительно отвечает за анализ данных.

Руководство финансами требует широкого спектра знаний и навыков. Помимо компетенций в области финансов и бухгалтерского учета эта роль обязывает быть бизнес-стратегом и разбираться в технологиях. Используя свои знания, финансовый директор формирует финансовую стратегию, направленную на рост организации и увеличение ее прибыльности, а также разрабатывает планы оптимизации финансовых активов.

Имеющиеся сегодня учебные программы и системы сертификации для подготовки финансовых директоров являются зрелыми и устоявшимися. Можно было бы ожидать наличия подобных образовательных стандартов и в отношении роли руководителя по работе с данными. Однако пока таких стандартов нет, как нет и единого мнения относительно того, какие виды сертификатов подходят для этой роли.

Таким образом, в то время как организации быстро осознают потребность в руководителях, которые сосредоточены исключительно на данных, на открывающиеся вакансии не так легко найти подходящих кандидатов. Часто поиск проблематичен, дорогостоящ и разочаровывает.

По мнению многих организаций, наиболее точно отражают их потребности в осуществлении руководства работой с данными эксперты в области информационных технологий. Это означает, что новый назначенец, скорее всего, будет склонен рассматривать возникающие в организации проблемы с данными

²³⁷ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

²³⁸ CMMI Institute. Data Management Maturity (DMM)SM. Build, improve, and measure your Enterprise Data Management Program. ISACA, 2021. – URL: <https://cmmiinstitute.com/dmm>.

²³⁹ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

как технические и требующие, соответственно, технических решений.

Пока организации не поймут, что извлечение максимальной ценности из их данных требует ориентированного на данные квалифицированного и ответственного руководителя, они не смогут изменить устоявшийся образ мышления в сторону большей дата-центричности.

3. Неспособность внедрить программный подход к обеспечению совместного использования данных

Тема управления корпоративными данными в рамках ИТ-проекта обычно обсуждается в контексте вопроса: «Каким образом деятельность по управлению данными вписывается в ИТ-проект?» Однако требования к данным организации развиваются, как правило, более медленными и устойчивыми темпами, отличающимися от темпов развития требований, связанных с ИТ-проектами. Например, определенная однажды область корпоративных данных, обеспечивающая бизнес-цель выхода на новый рынок, может в дальнейшем многие годы использоваться все новыми и новыми приложениями. Поэтому целесообразно рассматривать управление данными в контексте не отдельных ИТ-проектов, а всей их совокупности.

Выше, на рисунке 6.2 была показана взаимосвязь между разработкой архитектуры данных организации и ИТ-проектов. Сфера ИТ работает, как правило, в соответствии со стандартизированным методическим подходом, основанным на проектах, чтобы обеспечить контроль затрат и повысить производительность.

Развитие области совместно используемых данных не ориентировано на проектные методы, оно должно осуществляться на постоянной основе в соответствии с единой стратегией и требует целенаправленных и контролируемых усилий в масштабе всей организации. Общие данные должны проектироваться и документироваться отдельно от ИТ-проектов. Эта работа основополагающая, поскольку данные должны быть определены, прежде чем их можно будет использовать в любом индивидуальном ИТ-проекте.

Вопрос, прозвучавший в начале этого подраздела, должен быть поставлен по-другому: «Как ИТ-проекты вписываются в деятельность по управлению данными?»

Требования к данным, их форма и семантика продолжают развиваться по мере того, как бизнес реагирует на изменения окружающей среды, пока существует организация. Проекты же работают в соответствии с конечным промежутком времени. Было бы неправильно предполагать, что узконаправленные, рассчитанные по времени усилия в рамках конкретного проекта смогут быстро, точно и полностью определить информацию, необходимую всей организации для успешной работы. Кроме того, в одном проекте трудно разработать общепринятые бизнес-процессы и правила, которые позволят обмениваться данными внутри организации стандартным образом.

К сожалению, часто перед конкретными ИТ-проектами не ставятся задачи по определению данных, необходимых в масштабе всей организации. Когда это происходит, потребности отдельного проекта удовлетворяются в ущерб потребностям организации.

Чтобы эффективно использовать свои информационные активы, организации нужно сосредоточить усилия на углубленном понимании потребностей в данных, не отвлекаясь на индивидуальные ИТ-проекты. Необходимо, чтобы работы по выработке требований к корпоративным данным развивались отдельно и независимо от ИТ-проектов, опережая их. Такая деятельность должна быть оформлена в организации в виде специальной программы. Имеется в виду классический подход проектного менеджмента, в котором под программой понимается «совокупность взаимосвязанных проектов и другой деятельности, направленной на достижение общей цели и реализуемой в условиях общих ограничений»²⁴⁰. Программа позволяет решать проблемы, где один проект не сможет справиться. Основные отличия проекта от программы представлены в таблице 6.1.

4. Отсутствие согласованности программы в области данных с ИТ-проектами

Разработка и реализация корпоративных инициатив в области данных на уровне управления ИТ-проектами создает трудноразрешимые конфликты. У каждого проекта есть начало, середина и окончание. Однако, как мы уже обсуждали, данные организации должны быть исследованы, проанализированы,

240 См. ГОСТ Р 54871-2011 «Проектный менеджмент. Требования к управлению программой».

смоделированы и спроектированы на уровне выше отдельных ИТ-проектов. ИТ-проекты предназначены для реализации ИТ-решений. Выработка понимания того, какие данные необходимы организации для создания или поддержания наиболее выгодной конкурентной позиции, – это совершенно другое направление деятельности. Его должен возглавлять руководитель с опытом работы в бизнесе, отвечающий за управление информационными активами организации. Здесь существует аналогия с работой руководителей, которым поручено управлять другими видами организационных активов, такими как финансы, имущество и персонал, и которые выполняют эту работу на постоянной основе. Для достижения отдельных целей они могут создавать конкретные проекты, но фундаментальная направляющая и объединяющая все усилия работа продолжается без перерыва в течение всей жизни организации. При этом блок ИТ не должен иметь возможности что-либо делать с данными без одобрения бизнеса.

Таблица 6.1.

Отличия проекта от программы*

Признак	Проект	Программа
1. Объем понятия	Это ограниченное во времени намерение создать уникальный продукт или услугу	Это группа проектов, объединенных общей целью, управлением, ресурсами, миссией
	Вывод: программа — понятие более широкое, включающее в себя совокупность проектов, т. е. масштаб задач более обширен и требует большего взаимодействия, а проект — понятие более концентрированное и нацелено на конкретный результат — продукт или услугу	
2. Сложность	Проект — строго выстроенный план действий по решению конкретной сложной задачи, требующий взаимодействия только внутри ее	Программа более многогранна по масштабу, организации, реализации, управлению и требует значительного взаимодействия при решении отдельных задач (проектов)
	Вывод: реализация программы сложнее, чем выполнение проекта, так как все слагаемые задачи нуждаются в четкой координации	
3. Ограниченность	Сроки реализации проекта четко определены ввиду основных ограничений (время, ресурсы) и конкретных поставленных целей, т. е. проект имеет определенную точку начала и завершения	Сроки реализации программы широкие ввиду неясности последствий от совокупности отдельных проектов, т. е., скорее всего, нет определенной точки завершения
	Обеспеченность другими ресурсами (материалами, энергией, технологиями, трудовыми, информационными и т. д.) — проект четко выверен по количеству, объему и календарному использованию	Обеспеченность другими ресурсами (материалами, энергией, технологиями, трудовыми, информационными и т. д.) — у программы это носит общий характер по формированию, расходованию и по объемам
	Вывод: программа имеет более широкие сроки реализации, финансирование и применение других ресурсов носит общий характер по объему и расходованию, а у проекта — все строго по плану	
4. Технологичность	Проект по своей сути уникален и зачастую применяет инновации и в создании продукта, и в подходах к реализации и управлению	Программа в основном базируется на программном подходе к управлению и реализации
	Вывод: программа склонна к стандартизации ввиду сложности прогнозирования результата и ответственности, а проект — к новым технологиям ввиду большой конкурентоспособности	

Признак	Проект	Программа
5. Управление	Управление проектом — это использование знаний, опыта, методов, средств и технологий при выполнении проекта на всем протяжении его жизненного цикла с целью достижения определенных результатов и удовлетворенности ожиданий участников	Управление программой — это совокупность процессов управления несколькими взаимосвязанными проектами, направленных на повышение эффективности использования ресурсов, снижение рисков и успешное завершение каждого проекта и достижение основной цели
	Вывод: программа представляет собой бизнес-систему управления всеми проектами и связанными с ними мероприятиями для достижения необходимого эффекта с учетом стратегической инициативности принятия решений, а проект обладает только тактической инициативностью в реализации и получении конкретного результата в условиях ограничения	
6. Результат	Это конкретный продукт или услуга, согласно заявленной цели выполненная в установленный срок и в рамках выделенных ресурсов	Подразумевает качественное изменение состояния, вызванное реализацией запланированных задач
	Вывод: результат программы — это последствия (выгоды), а результат проекта — продукт или услуга	

* Вершинин В. П. Верификация отличий проекта от программы // Научно-практический журнал «Вестник Университета Российской академии образования», 1/2020: 108–116. – URL: <https://cyberleninka.ru/article/n/verifikatsiya-otlichiy-proekta-ot-programmy>.

Вид деятельности, которую осуществляют руководители по работе с данными, – это программа. Работа начинается и продолжается до тех пор, пока организация не решит, что ее больше не нужно выполнять. В рамках ее выполнения может одновременно поддерживаться несколько программ, проектов и конкретных мероприятий, которые составляют общую программу работы с данными.

Путаница, связанная с распределением ответственности в отношении данных, приводит к появлению все новых и новых осложнений, отнимающих много времени и ресурсов. Например, при попытке поиска первопричины низкого качества данных быстро приходит понимание, что существует фундаментально неправильное представление о том, кто на самом деле отвечает за решение проблемы. Многочисленные опросы представителей бизнеса показывают, что, по их мнению, качество данных обеспечивается блоком ИТ. В свою очередь представители ИТ-подразделений считают, что обеспечение качества данных – это функция, выполняемая бизнесом. За последние несколько десятилетий в организациях укрепилось представление о том, что данные являются проблемой ИТ и что именно ИТ-директор несет ответственность за ее решение. Между тем, поскольку ИТ-директора должны учитывать множество других вещей помимо данных, качество данных стало причиной разрыва между бизнесом и ИТ. Исследования показали, что только примерно 10 % всех организаций получают положительную отдачу от инвестиций в управление данными, при этом около 30 % сталкиваются с отрицательными результатами в отношении рентабельности инвестиций²⁴¹.

Очевидно, что подход организаций к управлению информационными активами должен кардинально отличаться от принятого в настоящее время. Необходимо создание новой руководящей должности на стороне бизнеса с правами и обязанностями по руководству работой с данными.

241 Там же.

После выработки общего понимания потребностей организации в данных они должны постоянно уточняться со стороны бизнеса в рамках самостоятельного направления деятельности, обеспечивающего предоставление четких требований в отношении данных для отдельных ИТ-проектов. Появление стандартов и спецификаций данных, доступных для каждого ИТ-проекта, позволит сократить объем работ и путаницу, с которой они регулярно сталкиваются. По мере созревания соответствующих процессов управления данными каждый новый проект по созданию приложений сможет предоставлять задействованным в процессах сотрудникам документированную обратную связь на основе сформированных подробных сведений о данных. Однако для этого организациям необходимо сначала пересмотреть свое понимание данных и применить это новое понимание к проектам и программам. Простое указание ИТ-специалистам изменить способ их работы не приведет к успеху. Изменения должны начинаться с высшего звена управления, что обеспечит использование преимуществ, создаваемых индивидуальными ИТ-проектами, в масштабах всей организации.

Чтобы проиллюстрировать затруднительное положение, в котором оказываются специалисты по управлению данными, сравним два подхода к выполнению ИТ-проектов.

Схема на рисунке 6.4 отражает традиционную, ориентированную на приложения практику проектирования информационных систем, которая принята сегодня в большинстве организаций.



Рис. 6.4.

Традиционная практика реализации ИТ-проектов*

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Следуя данному подходу, организация разрабатывает свою общую бизнес-стратегию, затем определяет конкретные ИТ-проекты, которые предположительно помогут ее реализовать, и наконец ИТ-проекты определяют, какие данные и информацию организация должна использовать, чтобы быть конкурентоспособной. Таким образом, планирование и проектирование данных, используемых организацией, осуществляется в ходе отдельных ИТ-проектов, после их запуска.

Описанный подход имеет очевидный недостаток – он не обеспечивает нужной согласованности с бизнес-стратегией. При такой модели бизнес-операции определяют, какие данные необходимы, хотя обычно происходит наоборот. Кроме того, ориентация на приложения приводит к тому, что данные тесно интегрированы с бизнес-процессами, реализуемыми программным обеспечением, что затрудняет их обслуживание, изменение и развитие. Требования к данным определяются требованиями к конкретному приложению, а не потребностями организации в целом. В результате лишь малая часть данных, обрабатываемых приложением, может быть использована повторно.

Часто ориентированная на приложения практика проектирования приводит к тому, что создаваемая в ходе ИТ-проекта информационная система превращается в так называемый информационный силос (silo²⁴²) – изолированную систему, которая не может обмениваться данными с другими системами^{243, 244}.

²⁴² В переводе с английского silo – «силосная яма (бункер)» или «силосная башня».

²⁴³ Бакланов И. Г. SILOсная яма современного телекома, или Блеск и нищета Open Source в России // ИнформКурьер-Связь (ИКС), 3/2019: 27–29. – <https://www.iksmmedia.ru/articles/5604674-SILOsnaya-yama-sovremennogo-telekom.html>.

²⁴⁴ Aiken P. The Seven Deadly Data Sins – Emerging from Management Purgatory. Dataversity, 2017. – URL: <https://www.dataversity.net/data-ed-slides-seven-deadly-data-sins-emerging-management-purgatory/>.

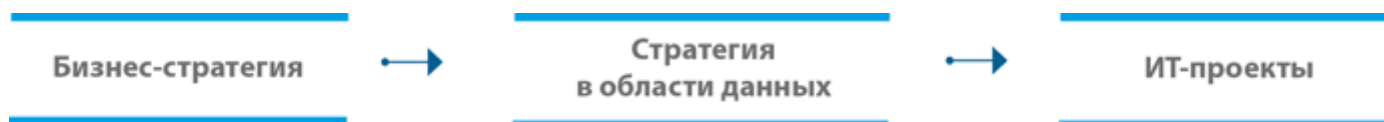


Рис. 6.5.

Дата-центричная практика выполнения ИТ-проектов*

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Как и при ориентации на приложения, при дата-центричном подходе организация прежде всего определяет свою общую бизнес-стратегию. Но далее (до начала инициации ИТ-проектов) ответственными за работу с данными со стороны бизнеса должны быть определены основные инициативы в области данных, необходимые для достижения стратегических целей. Должна быть выработана стратегия работы с данными в масштабах всей организации, предполагающая общую, совместно используемую архитектуру данных. В результате проведения указанных мероприятий ИТ-проекты получают необходимые определения и спецификации в отношении данных уже в начале своего выполнения.

При таком построении проектной работы блок ИТ не несет основного груза ответственности за проведение анализа потребностей в данных на уровне организации. Вместо этого он может сосредоточиться на технических решениях, которые в отношении данных уже приведены в соответствие со стратегическими целями организации. Совместное осуществление руководства в области данных и руководства в области ИТ показано на рисунке 6.6.

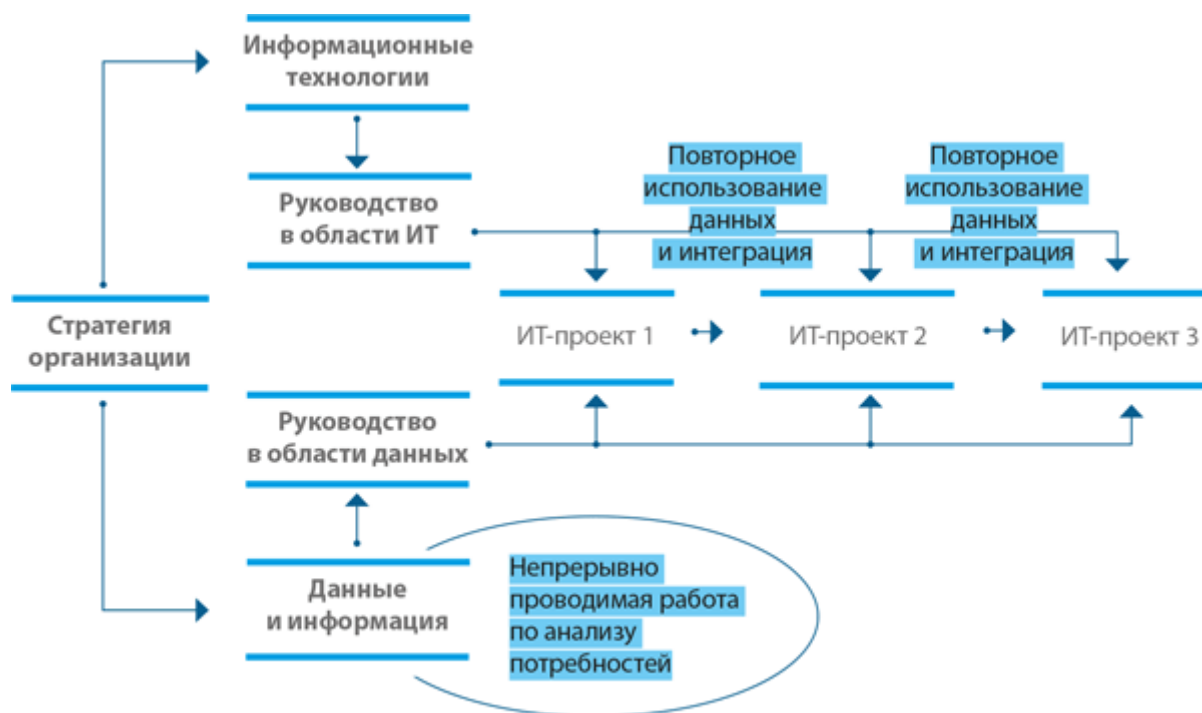


Рис. 6.6.

Совместное осуществление руководства в области данных и руководства ИТ*

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Дата-центричный подход предлагает множество явных преимуществ по сравнению с традиционным ИТ-подходом.

- Информационные активы планируются и развиваются как по-настоящему корпоративные.
- ИТ-проекты обеспечивают потребности в данных в масштабах всей организации.
- Иницируемые ИТ-проекты могут быть существенно оптимизированы с точки зрения сложности решений и привлекаемых ресурсов, поскольку они будут использовать существующую организационную

модель данных.

- Увеличивается совместное и повторное использование данных, а также снижается взаимозависимость данных и программного обеспечения.
- Расширяются возможности обмена данными, сокращается их дублирование и повышается качество.
- Упрощается сопровождение информационных систем, особенно когда данные совместно используются несколькими функциональными направлениями.
- Увеличивается совместное использование метаданных и обеспечивается более ясное и полное понимание бизнес- и технических требований.
- Повышается эффективность деятельности сотрудников, работающих с данными, и улучшается интеграция приложений с развивающимися бизнес-процессами и практиками.

В конце концов становится ясно, что внедрение дата-центричного подхода может привести к оптимизации корпоративного ИТ-ландшафта за счет создания более качественных и простых в обслуживании программных приложений и базовых информационно-технологических решений. Одновременно это обеспечивает более точную фокусировку работы с данными на достижении стратегических целей организации.

5. Неспособность адекватно управлять ожиданиями

Прежде чем организация сможет воспользоваться преимуществами дата-центричного подхода, она должна выполнить два условия:

- начать управлять ожиданиями;
- привести себя в соответствие с новой парадигмой.

Важно отметить, что оба условия необходимы для использования дата-центричной модели.

На достижение ощутимых изменений в поведении могут потребоваться годы. Ответственность за то, чтобы организация понимала, каким образом управление данными оказывает положительное влияние на ее способность достигать своих стратегических целей, лежит на руководителе, отвечающем за работу с данными. Во главе с ним организации необходимо развивать корпоративные компетенции в области управления данными в рамках серии соответствующих программ, проектов и мероприятий. На этом пути он должен устанавливать организационные ожидания и управлять ими. Это предполагает тщательную поддержку баланса между планируемыми изменениями и достижением реальных, ощутимых и измеримых результатов. В долгосрочной перспективе ситуация может быть очень неустойчивой.

Часто очень сложно бывает показать, каким образом данные поддерживают стратегию организации, и не менее сложно управлять ожиданиями в отношении новых инициатив по управлению данными. Если организация намерена использовать данные в качестве актива, она сначала должна понять их двойственную природу – либо она будет использовать данные в интересах реализации своей стратегии, либо те же самые данные станут препятствием. Нейтральной позиции не существует, поэтому организации необходимо контролировать факторы, которые могут помешать ей в полной мере использовать корпоративные информационные активы.

Высший менеджмент должен понимать и признавать, что на изменения потребуется время. Ответственному за данные руководителю необходимо четко сформулировать повестку дня, поддерживающую баланс между развитием конкретных способностей и измеримыми результатами. Она должна быть описана таким образом, чтобы другие могли оценить выгоды от предлагаемых инициатив. До тех пор пока организации придерживаются практики выполнения ИТ-проектов, ориентированной на приложения, ответственному за данные придется объяснять, почему разработка корпоративной архитектуры данных и в целом внедрение грамотного управления данными требуют времени для реализации на должном уровне. Кроме того, ему придется защищать результаты своей деятельности, пока не будет доказана рентабельность инвестиций.

Многие организации активно стремятся проводить работы в области больших данных и расширенной аналитики. Однако они должны отчетливо осознавать, что если действительно хотят доверять получаемым при этом результатам, то должны осуществлять учет своих данных и управление ими на протяжении всего их жизненного цикла, от получения до ликвидации. Единственный способ добиться этого – обеспечить, чтобы специалисты по управлению данными и ИТ-эксперты работали вместе как команда. По мере того

как их способности по управлению данными становятся все более зрелыми, они должны фиксировать полученные знания с целью передачи их своим последователям.

Организации должны отдавать себе отчет, что они не могут просто купить способности по эффективному управлению данными, как бы заманчиво это ни было. Чтобы быть устойчивыми, эти навыки должны целенаправленно развиваться с течением времени. Только понимая свои текущие способности, а также свои сильные и слабые стороны, можно надеяться на успех.

6. Отсутствие последовательности в реализации стратегии в области данных

В большинстве организаций команды по управлению данными не готовы одновременно фокусировать свои усилия и на улучшении операционной деятельности организации, и на развитии инноваций. Практики управления данными развиваются постепенно, и по мере роста их зрелости фокус стратегии организации в области данных последовательно перемещается по четырем квадрантам, представленным на рисунке 6.7. Изменение фокуса стратегии условно отражено в виде четырех уровней ее развития – как бы четырех ее версий: от V1 до V4.

В организациях, соответствующих уровню V1, данные не считаются стратегически важными с точки зрения обеспечения стратегии бизнеса. Управление данными осуществляется в основном в рамках локальных рабочих групп и не выходит за их пределы. Организации не управляют своими данными на корпоративном уровне и не рассматривают их как стратегический актив. Вместо этого прилагаются минимальные усилия по обслуживанию данных с целью обеспечения текущей операционной деятельности. Образно говоря, организации сосредотачивают свои усилия на отчетности об остатках наличности, а не на развитии способностей к прогнозированию наличности.

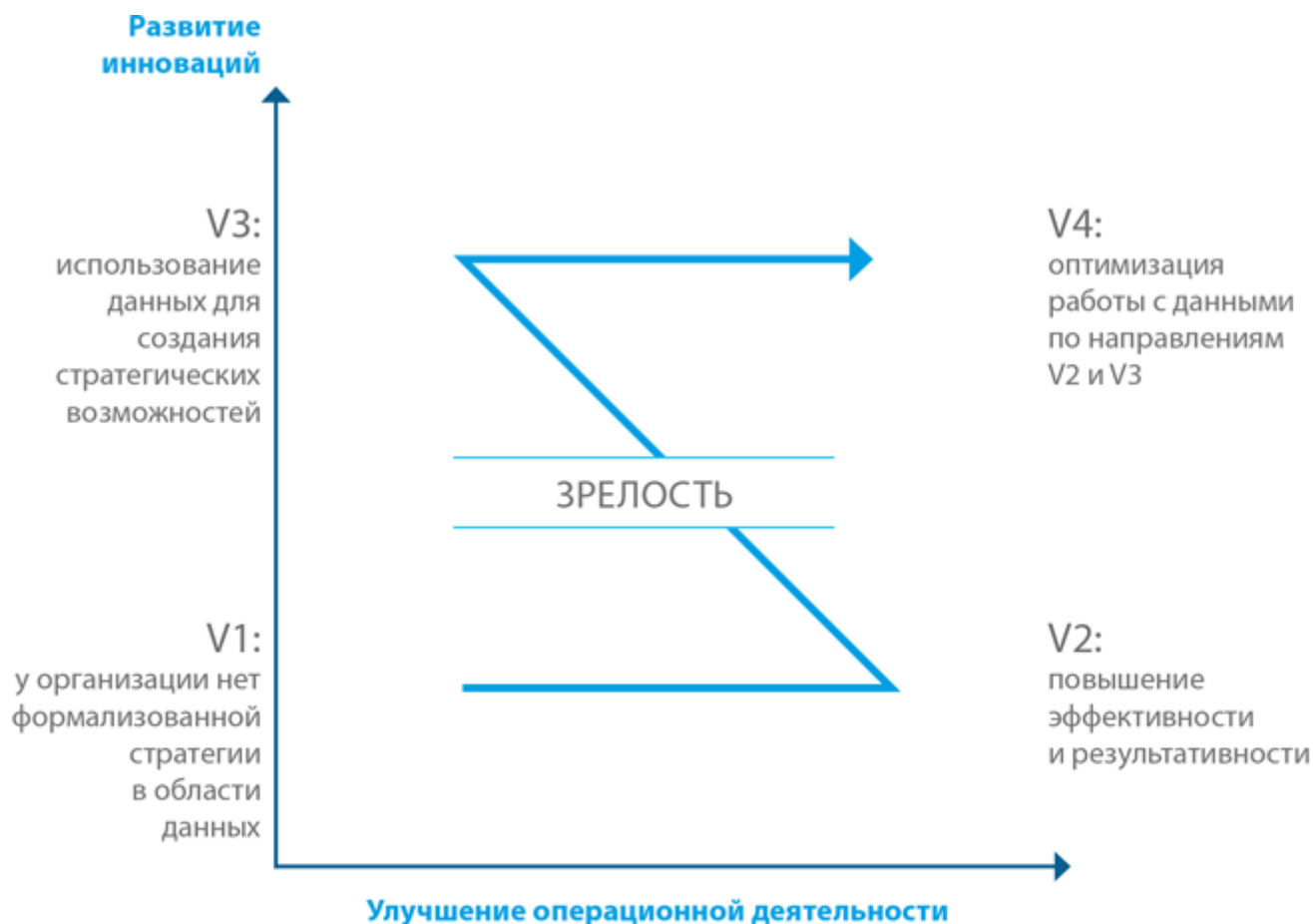


Рис. 6.7.

Смена фокуса стратегии в области данных по мере роста зрелости практики управления данными*, **

* Aiken P. The Seven Deadly Data Sins – Emerging from Management Purgatory. Dataversity, 2017. – URL: <https://www.dataversity.net/data-ed-slides-seven-deadly-data-sins-emerging-management-purgatory/>.

** Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Организации с уровнем V2 следуют стратегии работы с данными, ориентированной в первую очередь на повышение эффективности и результативности операционной деятельности. Например, на оптимизацию управления цепочками поставок.

Организации, реализующие стратегию в версии V3, осознали, что данные могут помочь им переосмыслить себя и занять лучшие позиции на рынке. Например, они могут проводить работу по анализу данных и выявлению перспективных категорий клиентов с целью создания новых видов продуктов и услуг.

Деятельность по работе с данными на уровне V3 часто сопряжена с серьезными затратами и имеет не очень высокую отдачу от сделанных инвестиций. Это характерно, в частности, для многих проектов в области больших данных и науки о данных.

Наконец, организации с уровнем V4 – это те, которые успешно освоили методы управления данными на уровнях V2 и V3.

Важно понимать, что большинству организаций для эффективного осуществления стратегии в области данных требуется серьезная перестройка. Поэтому руководители по работе с данными должны ориентировать свою стратегию прежде всего на обеспечение отдачи от уже сделанных организацией долгосрочных инвестиций, включая вложения в обучение и образование. И только потом – на поддержку

инновационной деятельности.

7. Отсутствие внимания к проблемам, связанным с культурой и управлением изменениями

Джон Лэдли выделяет несколько основных проблемных областей, недостаточное внимание к которым может способствовать срыву внедрения в организации программы руководства работой с данными. Они перекликаются с уже перечисленными «смертными грехами»²⁴⁵:

- **Грамотность в области данных (data literacy).** Отсутствие понимания важности данных.
- **Выравнивание с бизнесом и приоритетность.** Руководство в области данных должно поддерживать потребности бизнеса и входить в число приоритетных и развивающихся направлений деятельности организации.
- **Качество данных.** Организация не может рассчитывать на получение ожидаемой отдачи от данных при их недостаточном качестве.
- **Обучение и коммуникации.** Основными компонентами внедрения стратегии в области данных являются постоянное обучение и налаженные коммуникации.
- **Заблуждение «сначала стюарды».** Во многих организациях распространено ошибочное мнение, что, пока не выработано поддерживаемое руководством ясное представление о задачах и функциях корпоративного управления данными, необходимо делать хоть что-то. Поэтому нужно начинать с назначения ответственных за работу с данными на местах – дата-стюардов (data stewards). Однако при таком подходе для данной категории сотрудников трудно сформировать четкие требования к выполняемым обязанностям и определить зоны ответственности.
- **Заблуждение «сначала технологии».** Этому распространенному заблуждению противостоит утверждение о ценности № 2 из рассмотренной выше Доктрины в области данных. Тут можно только повторить, что без четкого понимания потребностей в информационном обеспечении деятельности организации приобретаемые технологии могут привести даже к снижению ее эффективности.
- **Отнесение проблематики работы с данными исключительно к сфере ИТ.** В большинстве организаций распространено ошибочное представление: «Данные – это ответственность ИТ».
- **Понимание руководства в области данных как проекта.** Как уже отмечалось выше, вид деятельности, которую осуществляют руководители по работе с данными, – это не отдельный проект с фиксированным сроком завершения, а программа. Работа начинается и продолжается до тех пор, пока организация не решит, что ее больше не нужно выполнять.
- **Ориентация на наем специалистов извне.** Во многих компаниях и учреждениях предпочитают внедрять деятельность по руководству в области данных не за счет использования внутренних ресурсов, а приглашая нужных специалистов извне. Такая практика препятствует развитию необходимых компетенций внутри самой организации.
- **Управление изменениями.** Успех внедрения в организации руководства в области данных зависит от ее способности проводить необходимые изменения.

Хотя управление изменениями и стоит в перечисленном списке на последнем месте, Джон Лэдли выделяет эту область в качестве ключевой. Со всеми остальными проблемными областями можно справиться, грамотно проводя организационные изменения (рис. 6.8).

²⁴⁵ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

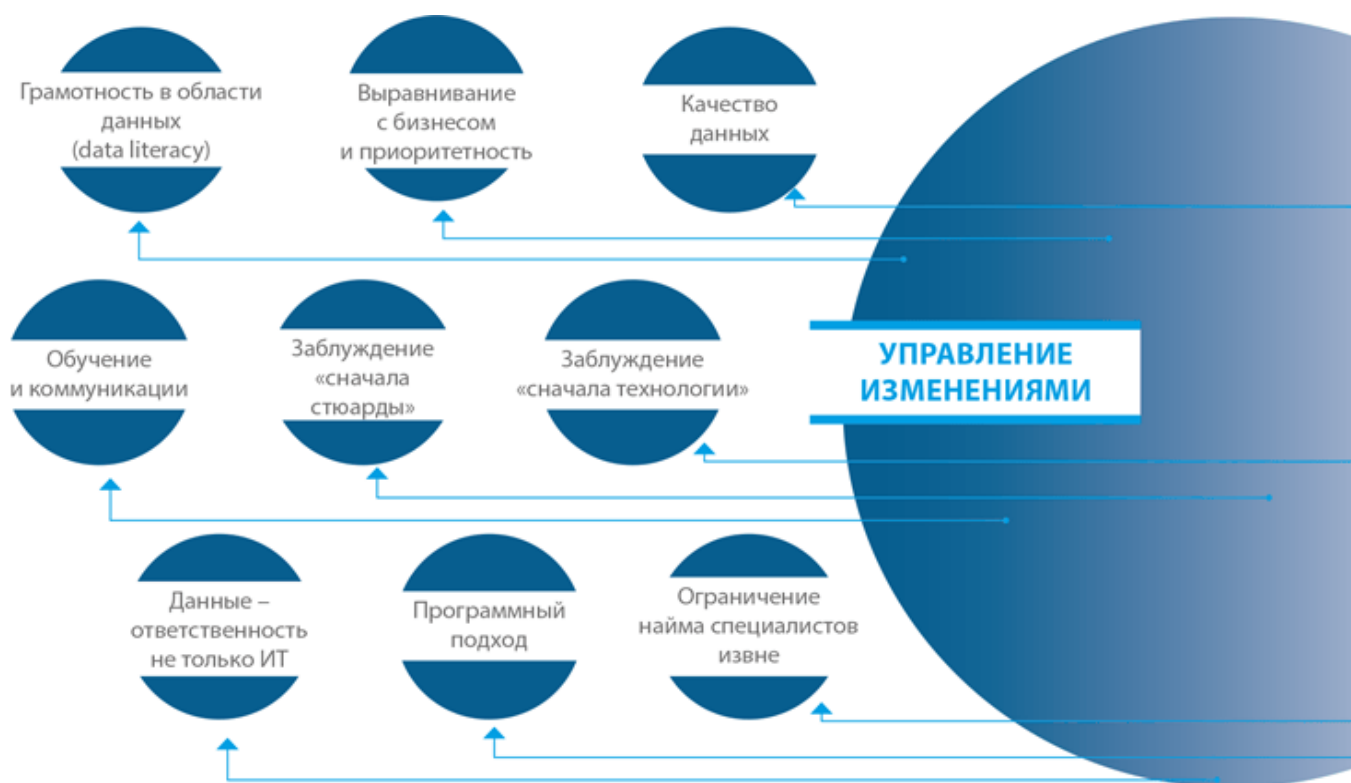


Рис. 6.8.

Управление изменениями как ключевой элемент в решении проблем работы с данными

Не случайно методам управления организационными изменениями (Organizational Change Management, OCM) посвящена отдельная глава DAMA-DMBOK²⁴⁶. Проводимые изменения должны привести к постепенному изменению организационной культуры и в итоге к изменению поведения каждого сотрудника. От того, насколько эффективно будет продвигаться этот процесс, зависит успешность формирования и выполнения стратегии организации в области данных. Как заметил Питер Друкер²⁴⁷, «культура съедает стратегию на завтрак»²⁴⁸.

Литература к главе 6

- ГОСТ Р 54871-2011 «Проектный менеджмент. Требования к управлению программой».
- Вершинин В. П. Верификация отличий проекта от программы // Научно-практический журнал «Вестник Университета Российской академии образования», 1/2020: 108–116. – URL: <https://cyberleninka.ru/article/n/verifikatsiya-otlichiy-proekta-ot-programmy>.
- Evans N., Price J. Enterprise information asset management: the roles and responsibilities of executive boards // Knowledge Management Research & Practice, 2015, 14: 353–361. – URL: <https://www.experiematters.com.au/wp-content/uploads/2018/03/Enterpriseinformation-asset-management-the->

²⁴⁶ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

²⁴⁷ Питер Друкер (Peter Ferdinand Drucker) – выдающийся американский экономист, публицист, педагог, один из самых влиятельных теоретиков менеджмента XX века.

²⁴⁸ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

roles-and-responsibilities-of-executive-boards.pdf.

Глава 7. Построение цепочек

7.1. Цепочка ценности данных

В предыдущих главах мы говорили о том, что необходимое условие преодоления разрыва между данными и информацией – изменение отношения организации к своим данным. Она должна перестать относиться к данным как к побочному продукту операционной деятельности и начать рассматривать их как стратегически важный актив. Теперь обсудим две важнейшие концепции, направленные непосредственно на устранение указанного разрыва.

При обсуждении процессов управления данными часто рассматривают две цепочки – цепочку ценности данных (data value chain) и цепочку поставок данных (data supply chain). Что это за цепочки и как они соотносятся друг с другом? Рассмотрим сначала цепочку ценности данных.

Концепция цепочки ценности (value chain)²⁴⁹ была предложена Майклом Портером в 1985 году в книге «Конкурентное преимущество. Как достичь высокого результата и обеспечить его устойчивость»²⁵⁰. Цепочка ценности – это система взаимозависимых видов деятельности по выпуску компанией на рынок своих товаров и услуг. Эти виды деятельности направлены на создание потребительской ценности, благодаря которой покупатель приобретает товар или услугу. Они объединены связями, которые возникают, когда способ выполнения одного вида деятельности влияет на стоимость или эффективность других видов деятельности²⁵¹.

Когда все виды деятельности (процессы) систематизированы и схематически представлены в виде единой цепочки, становятся видны конкурентные преимущества продукта и возможности для улучшения бизнес-процессов. Анализ цепочки ценности дает представление о том, как устроен бизнес.

249 Часто цепочка ценности также называется «цепочкой создания ценности» или «цепочкой создания стоимости».

250 Porter M. Competitive Advantage. Creating and Sustaining Superior Performance. Free Press, New York. 1985.

251 Porter M., Millar V. How information gives you competitive advantage // Harvard Business Review, 1985. 63 (4): 149–160. – URL: <https://palfreymanventures.net/porter-and-millar-1985.pdf>.



Рис. 7.1.

Виды деятельности (процессы), составляющие цепочку ценности*

* Swiss Business Hub Russia. Анализ цепочки создания стоимости: пять вопросов и ответов. Swiss Business Hub Russia, 2019 (совместно с газетой «Ведомости»). – URL: <http://swissbusiness.vedomosti.ru/page/qa>.

У любого бизнеса есть возможность оптимизировать существующие процессы или выстроить новые. Анализ цепочки ценности помогает:

- сформировать бизнес-стратегию;
- найти возможности для оптимизации бизнес-процессов или выстроить новые;
- увидеть конкурентные преимущества продукта;
- сократить затраты.

Согласно концепции Портера, всю работу компании можно разделить на основные и вспомогательные виды деятельности (процессы) (рис. 7.1). Каждый вид деятельности можно декомпозировать до уникальных, присущих только данному бизнесу действий – звеньев цепочки ценности^{252, 253}.

Основные виды деятельности связаны с физическим созданием продукта, его маркетингом и доставкой покупателям, а также его поддержкой и обслуживанием после продажи.

Вспомогательные мероприятия обеспечивают ресурсы и инфраструктуру, которые позволяют осуществлять основные виды деятельности. В каждом основном виде деятельности используются приобретенные ресурсы, человеческие ресурсы и сочетание технологий. Управление инфраструктурой компании, включая общее управление, юридическую работу и бухгалтерский учет, поддерживает всю цепочку.

В рамках каждого из этих общих видов деятельности компания будет выполнять различные работы в зависимости от конкретного бизнеса. Например, сервисное обслуживание часто включает в себя такие действия, как установка, ремонт, настройка, модернизация и управление резервом запчастей.

На основании анализа цепочки ценности выявляются ключевые драйверы бизнеса – сильные стороны

²⁵² Репин В. Описание и анализ бизнес-процессов: цепочки ценности и Work Flow. Корпоративный менеджмент. 2007. – URL: https://www.cfin.ru/itm/bpr/comparing_analysis.shtml.

²⁵³ Swiss Business Hub Russia. Анализ цепочки создания стоимости: пять вопросов и ответов. Swiss Business Hub Russia, 2019 (совместно с газетой «Ведомости»). – URL: <http://swissbusiness.vedomosti.ru/page/qa>.

компания, которые могут принести наибольшую прибыль. Цель такого анализа – понять, как увеличить конкурентное преимущество продукта и компании в целом.

При анализе цепочки ценности можно обнаружить узкие места (разрывы), которые возникают обычно в смежных подразделениях компании. Например, подразделение, отвечающее за продажи, вместо непосредственных продаж вынуждено заниматься непрофильной деятельностью (вопросами бухгалтерии, склада, доставки и т. д.). Несогласованность действий различных звеньев цепочки отнимает время у участников процесса и негативно влияет на компанию в целом.

С учетом возможности подобных разрывов анализу подлежат все бизнес-процессы, объединения всех звеньев. Таким путем выявляют трудные участки и способы решения связанных с ними проблем.

Связи в цепочке часто создают противоречия, которые следует оптимизировать. Оптимизация может потребовать компромиссов. Например, более дорогостоящий дизайн продукта и более дорогое сырье могут снизить затраты на послепродажное обслуживание. Компания должна разрешать такие конфликты в соответствии со своей стратегией с целью достижения конкурентного преимущества.

Концепция цепочки ценности вполне применима по отношению к данным. В многочисленных источниках можно найти различные, но в целом примерно схожие варианты рамочных структур (frameworks) цепочки ценности данных для различных видов организаций и отраслей^{254, 255, 256}.

Рассмотрим звенья типичной рамочной структуры цепочки ценности данных, ориентированной на крупные компании розничной торговли (рис. 7.2).

1. Сбор данных

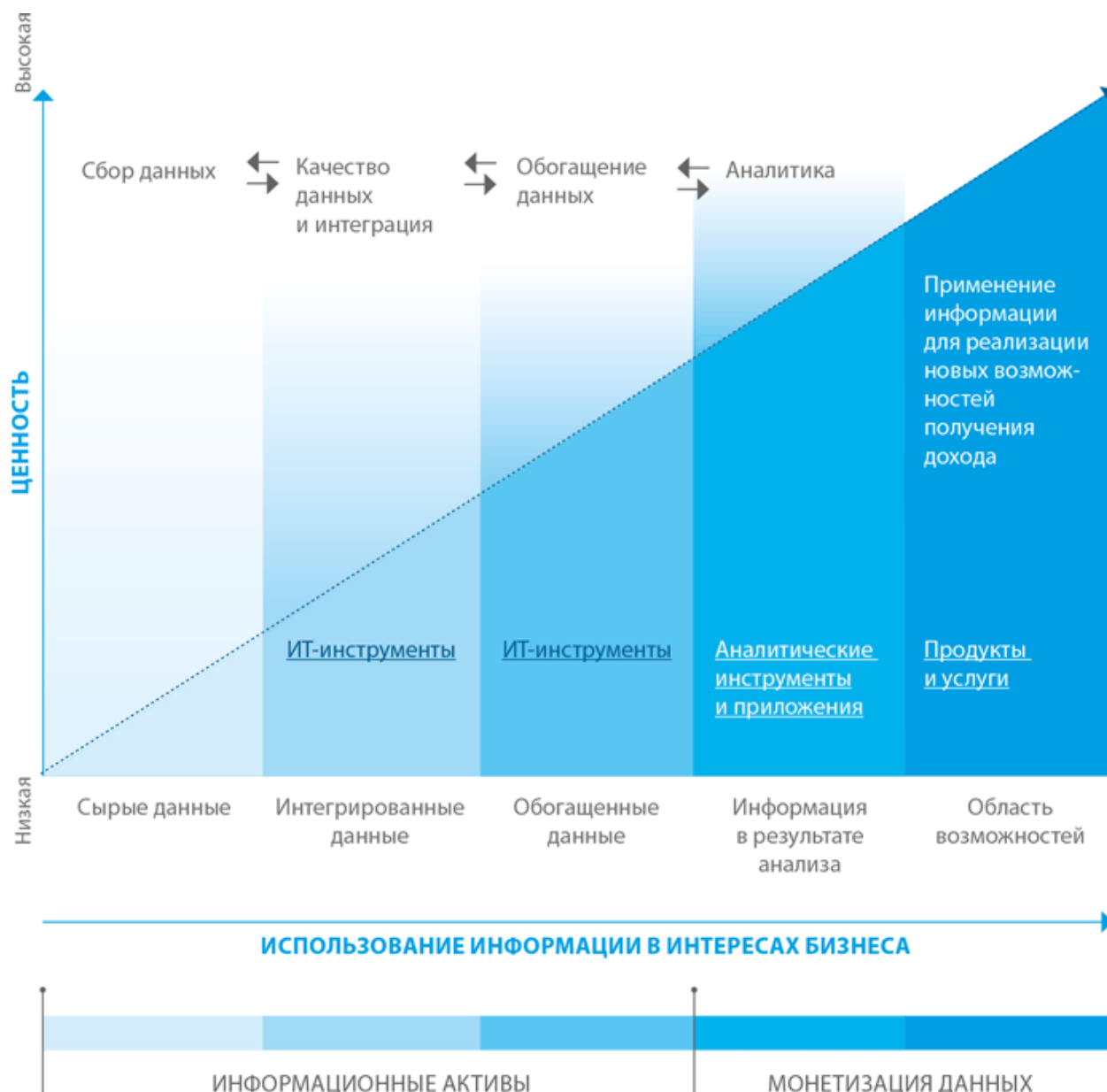
Существует множество современных эффективных методов сбора и хранения данных из различных источников (в том числе потоковых). Ключевая проблема сбора – разнообразие и объем данных. Это ставит вопрос о том, как хранить эти данные, но что более важно, как обработать их для хранения таким образом, чтобы они были готовы к анализу. Большинство организаций просто помещают собранные сведения в озера данных (data lakes), но, чтобы подготовить такие данные для анализа, часто требуются месяцы. С учетом того что пригодность данных со временем падает (см. главу 5, раздел 5.3), возможности извлечения из них ценности могут к моменту анализа просто исчезнуть, если не провести подготовку сразу после сбора²⁵⁷.

²⁵⁴ Anderson C. *Creating a Data-Driven Organization*. O'Reilly Media, 2017. (Русский перевод: Андерсон К. Аналитическая культура. От сбора данных до бизнес-результатов. – М.: Манн, Иванов и Фербер, 2017.)

²⁵⁵ GSMA. *The Data Value Chain*. GSM Association – A. T. Kearney, 2018. – URL: https://www.gsma.com/publicpolicy/wp-content/uploads/2018/06/GSMA_Data_Value_Chain_June_2018.pdf.

²⁵⁶ Narang K. *Data Value Chain*. Agility Exchange (agilityexchange.com), 2020. – URL: <https://agilityexchange.com/data-value-chains/>.

²⁵⁷ Более подробно вопросы сбора и обработки больших объемов данных рассматриваются в главе «Обеспечение доступности и обслуживание данных: развитие» (глава 14).

**Рис. 7.2.**

Цепочка ценности данных*

* Kotorov R. The Data Value Chain: Steps for Monetizing Your Data. Integration Developer News, 2017. – URL: <https://www.idevnews.com/stories/6998/The-Data-Value-Chain-Steps-for-Monetizing-Your-Data>.

2. Качество и интеграция данных

После сбора данные нужно организовать таким образом, чтобы все элементы были собраны в значимые единицы (массивы данных). Компании взаимодействуют с клиентами с использованием различных каналов: лично, онлайн, через социальные сети и т. д. Наличие многих точек взаимодействия с одним заказчиком требует, чтобы все данные были собраны в единое представление или «золотую запись» о клиенте. Только тогда можно осуществить точную сегментацию, увеличить перекрестные продажи и продажи в целом. Ключевыми здесь выступают процессы управления основными данными (Master Data Management, MDM) и обеспечения качества данных, что обусловлено необходимостью более быстрого и

точного принятия решений. Неверные данные приводят к неправильным решениям. Неуправляемые данные часто становятся неиспользуемыми или неправильно используемыми данными²⁵⁸.

3. Обогащение данных

Часто организации получают внешние данные и добавляют их в свои собственные записи. Например, в розничной торговле добавление психографических данных – данных о предпочтениях и чертах характера – часто полезно для лучшего понимания клиента. В сочетании с другими данными они дают маркетологам лучшее представление о мотивации клиентов, позволяя им разрабатывать более эффективные программы и предложения.

Описанные первые три шага (звена цепочки ценности) создают единицы (массивы) данных, готовые для анализа. Чем они полнее, тем больше ценной информации можно из них извлечь.

4. Аналитика

Как только единицы данных будут созданы, организация сможет извлекать информацию, позволяющую понять, что произошло в прошлом и что можно сделать в будущем. Анализ может выявить тенденции, знание которых будет способствовать выявлению скрытых издержек или выявлению новых возможностей для продаж и получения дохода. Все чаще организации используют обогащение и анализ данных, чтобы понять, почему происходят те или иные события или почему потребители ведут себя определенным образом, и с помощью этих знаний они создают новые источники дохода²⁵⁹.

5. Монетизация

После завершения анализа у организации появляются информационные активы, которые могут быть дополнительно монетизированы. Полученная новая информация расширяет понимание того, как эти активы могут быть правильно использованы, и позволяет открывать новые возможности. Но само по себе понимание недостаточно для монетизации. Открытая возможность должна быть каким-то образом выведена на рынок – либо путем ее реализации, либо путем предоставления к ней доступа внешних сторон.

Таким образом, цепочка ценности данных – важный инструмент стратегического анализа. Ее построение и визуализация может помочь выявить разрывы, определить способы их устранения и повышения ценности данных, а также сформировать видение будущего состояния.

Применительно к промышленным и торговым компаниям основной метод повышения эффективности цепочки ценности производимой и поставляемой продукции – управление цепями поставок (Supply Chain Management, SCM)²⁶⁰.

В отношении вопросов повышения эффективности цепочки ценности данных также полезно применение подходов SCM. Поэтому в следующем разделе мы рассмотрим такое понятие, как цепочка поставок данных.

7.2. Цепочка поставок данных

Управление цепями поставок

Впервые термин «управление цепями поставок»²⁶¹ был использован в начале 1980-х годов компанией

258 Вопросам интеграции и управления основными данными посвящены главы «Обеспечение доступности и обслуживание данных: основы» (глава 12) и «Управление основными данными: практика внедрения» (глава 13). Управление качеством данных обсуждается в главе «Базовая поддержка жизненного цикла данных» (глава 15).

259 Применение аналитики рассматривается в главе «Использование данных и расширение возможностей применения» (глава 17).

260 Иванов Д. А. Управление цепями поставок. – СПб.: Изд-во Политехнического университета, 2009.

261 Следует заметить, что когда обсуждаются вопросы управления цепями поставок, то слово chain обычно переводится как «цепь», в то же время применительно к теме цепочек ценности чаще употребляется слово «цепочка».

i2 Technologies и консалтинговой фирмой «Артур Андерсен». А в 1982 году в Великобритании публикуется статья Оливера (R. Oliver) и Вебера (M. Webber) «Управления цепями поставок: логистика вырастает до уровня стратегии»²⁶², ставшая основой для разработки концепции SCM.

В наши дни концепция управления цепями поставок, широко используемая в странах с развитой промышленностью, представляет собой действенный способ увеличения прибыли и доли рынка. Зачастую подобный системный подход к менеджменту предприятия на всех уровнях воспринимается бизнесменами в качестве новой идеологии ведения дел. Существуют международные компетентные организации, которые специализируются на развитии стратегических преимуществ логистики, – Европейская логистическая ассоциация и Совет профессионалов в области управления цепями поставок²⁶³.

Управление цепями поставок с практической точки зрения – это системный подход к интегрированному планированию и управлению всем потоком информации, материалов и услуг от конечного потребителя через предприятия и склады до поставщиков сырья.

SCM – целостная концепция ведения бизнеса, объединяющая передовые организационные принципы и возможности современных информационных технологий. В основе SCM лежит понятие цепи поставок. Единого общепринятого определения цепи поставок не существует. Это понятие определяется по-разному различными авторами, но суть при этом остается, как правило, одинаковой.

Цепь поставок в объектном понимании – это совокупность организаций (предприятий-изготовителей, складов, дистрибьюторов, поставщиков услуг, экспедиторов оптовой и розничной торговли), взаимодействующих в материальных, финансовых и информационных потоках, а также потоках услуг от источников исходного сырья до конечного потребителя.

Цепь поставок в процессном понимании – это совокупность потоков и соответствующих им кооперационных и координационных процессов между различными участниками цепочки ценности²⁶⁴ для удовлетворения требований потребителей в товарах и услугах²⁶⁵.

Рационализация цепи поставок позволяет кардинально повысить эффективность цепочки ценности путем снижения совокупных затрат в звеньях цепи поставок.

Начиная примерно с 1990-х годов в связи с интенсивным развитием технологий хранилищ данных (data warehouse) и по мере формирования представления о данных как об активе специалисты в области управления данными стали продвигать идею цепочки поставок данных (Data Supply Chain, DSC)²⁶⁶ как

²⁶² Oliver R., Webber M. Supply-chain management: logistics catches up with strategy // Outlook, 1982. 5 (1): 42–47.

²⁶³ Коммерческий директор. Управление цепями поставок: методы, программные продукты и проблемы // Коммерческий директор, 2018. – URL: <https://www.kom-dir.ru/article/2136-upravlenie-tsepyami-postavok>.

²⁶⁴ Ключевая роль цепи поставок в создании цепочки ценности здесь вполне очевидна.

²⁶⁵ Там же.

²⁶⁶ Несмотря на то что применительно к теме управления цепями поставок слово chain обычно переводится как «цепь», при обсуждении поставок данных чаще употребляется слово «цепочка».

разновидности цепи поставок^{267, 268, 269, 270, 271}.

Цепочка поставок – это удобная метафора для визуализации, определения, уточнения и оценки процессов и ресурсов, поддерживающих жизненный цикл данных. Как мы упоминали в главе 4, она позволяет улучшить взаимодействие между отдельными функциями управления данными. Несмотря на то что термин «цепочка поставок» звучит так, будто фокусируется на поставщике, на самом деле цепочки поставок разрабатываются с учетом потребностей клиента. Таким образом, концепция может помочь специалистам по управлению данными сформировать более полное представление о бизнес-результатах, которые достигаются за счет использования информационных активов. Ориентация на результат позволяет более осознанно подойти к проблеме преодоления разрыва между данными и информацией, о котором мы говорили в главе 2^{272, 273}.

Даглас Лейни в книге «Инфономика: информация как актив: монетизация, оценка, управление»²⁷⁴ рассматривает возможность применения общепризнанных методических подходов в области SCM к управлению цепочками поставок данных²⁷⁵. Обсудим этот вопрос подробнее.

SCOR-модель

Признанная в мире референтная модель управления цепями поставок – SCOR-модель (Supply Chain Operations Reference model), референтная модель операций в цепях поставок. Специалисты многих западных компаний рассматривают ее как международный межотраслевой стандарт управления цепями поставок.

SCOR-модель разработана американской консалтинговой компанией PRTM в 1996 году. Она была одобрена и систематически обновляется Советом по цепям поставок (SCC²⁷⁶), который в 2014 году

267 Вопросы управления поставками данных и поставщиками данных мы уже коротко обсуждали в главе 4.

268 Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

269 Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

270 Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

271 Spanaki K., Gurguc Z., Adams R., Mulligan C. Data Supply Chain (DSC): Research Synthesis and Future Directions // International Journal of Production Research, 2018. 56(3). DOI: 10.1080/00207543.2017.1399222 _ URL: https://dspace.lib.cranfield.ac.uk/bitstream/handle/1826/12800/Data_supply_chain-2017.pdf?sequence=4&isAllowed=y.

272 Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

273 Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

274 Там же.

275 Некоторые авторы, в частности Лейни, говорят о «цепочке поставок информации» (Information Supply Chain, ISC). В этой книге мы будем использовать чаще встречающийся термин «цепочка поставок данных».

276 Совет по цепям поставок (Supply Chain Council, SCC) – созданная в 1996 году независимая некоммерческая организация, объединяющая ведущие компании мира, среди которых производственные и торговые предприятия, логистические провайдеры, ритейлеры, разработчики программного обеспечения.

объединился с Американским обществом по управлению запасами и производством (APICS²⁷⁷). Модель входит в состав поддерживаемого APICS набора референтных рамочных структур (frameworks), описывающих важнейшие элементы цепочки ценности. Также в него включены референтные модели операций жизненного цикла продукта (Product Life Cycle Operations Reference model, PLCOR), операций в цепях проектирования (Design Chain Operations Reference model, DCOR) и ряд других, которые дополняют SCOR-модель в части специфических аспектов бизнес-деятельности. Текущая, двенадцатая версия SCOR-модели выпущена в 2017 году²⁷⁸.

SCOR-модель сочетает в себе три популярные управленческие концепции: реинжиниринг бизнес-процессов (business process reengineering), бенчмаркинг (benchmarking) и использование лучших практик (best practice). Описывающий модель документ состоит из четырех основных разделов.

1. Эффективность. В разделе более 250 метрик, которые представляют систему измеряемых показателей, позволяющих оценивать эффективность операций в цепях поставок (бенчмаркинг).

2. Процессы. Раздел содержит стандартные описания бизнес-процессов, выполняющихся в рамках цепи поставок, и взаимосвязей между ними.

3. Практики. Здесь описаны лучшие практики, которые обеспечивают значительное повышение эффективности процессов. Практика – это уникальный способ настройки процесса или совокупности процессов. Уникальность может быть связана с автоматизацией процесса, задействованными технологиями или специальными навыками, с необычной последовательностью выполнения операций или с особым подходом к распределению процессов между организациями и их взаимодействию.

4. Люди. В разделе приводятся стандартные определения навыков, необходимых для выполнения процессов цепи поставок.

В основе рамочной структуры SCOR лежат шесть ключевых процессов верхнего уровня (всего предусмотрено три уровня детализации).

1. Планировать (Plan). Процесс объединяет и координирует деятельность всех участников цепи поставок. Это интегрирующий элемент SCOR-модели, который обеспечивает баланс спроса и предложения. В рамках процесса определяются источники поставок, расставляются приоритеты потребительского спроса, планируются запасы, устанавливаются требования к системе дистрибуции, объемы производства, поставок сырья, материалов и готовой продукции, решается задача Make or Buy – производить самостоятельно или покупать. Также принимаются решения, которые относятся ко всем видам планирования ресурсов, мощностей и к управлению жизненным циклом товара.

2. Снабжать (Source). Процесс определяет деятельность по снабжению производителей всеми компонентами и услугами, необходимыми для создания товара или для его продажи. Устанавливаются такие процедуры, как оценка и выбор поставщиков, проверка качества поставок, заключение контрактов с поставщиками. Также с этим процессом связаны все процедуры, относящиеся к получению материалов – к их приобретению, транспортировке, входному контролю, приемке на хранение и т. д.

3. Делать (Make). Процесс включает операции, связанные с производством товара (физического продукта или услуги), и определяет специфические процедуры: производственные технологии, контроль качества, упаковка, хранение и выпуск (внутрипроизводственная логистика). К его структурным элементам относятся конструктивные и технологические изменения, управление производственными мощностями (оборудованием, зданиями и т. п.), производственные циклы, качество производства, график производственных смен и т. д.

4. Доставлять (Deliver). Весь цикл операций по доставке товара потребителям (собственными силами либо с привлечением логистических посредников) – от формирования заказа до получения

²⁷⁷ Американское общество по управлению запасами и производством (American Production and Inventory Control Society, APICS), в настоящее время известное как Ассоциация по управлению цепями поставок (Association for Supply Chain Management, ASCM), – крупнейшая некоммерческая ассоциация по продвижению передовых практик в области управления цепями поставок (<https://www.ascm.org>). Была основана в США в 1957 году и позже стала международной.

²⁷⁸ <http://www.apics.org/docs/default-source/scor-training/scor-v12-0-framework-introduction.pdf?sfvrsn=2>.

денежных средств.

5. Возвращать (Return). Деятельность по управлению возвратными материальными потоками (бракованная продукция, оборотная тара, различные виды отходов и т. д.).

6. Предоставлять возможность (Enable). Обеспечивающие или вспомогательные процессы, связанные с управлением цепями поставок. Сюда относятся управление бизнес-правилами, управление эффективностью, управление данными, управление ресурсами, управление инфраструктурой, управление контрактами, управление сетью цепей поставок, управление соблюдением нормативных требований, управление рисками и управление закупками в цепи поставок.

Для оценки процессов с помощью метрик в SCOR-модели выделены пять стратегических характеристик эффективности (Performance Attributes). Они используются для определения приоритетов и согласования эффективности цепи поставок с бизнес-стратегией организации. Все метрики распределены по пяти группам, соответствующим характеристикам эффективности.

1. Надежность (Reliability). Способность выполнять задачи должным образом. Надежность фокусируется на предсказуемости результатов процесса. Типичные метрики надежности включают: своевременность, надлежащее количество, надлежащее качество.

2. Оперативность (Responsiveness). Скорость выполнения задач – скорость, с которой цепь поставок поставяет продукцию заказчику. К примерам метрик относятся показатели времени цикла поставки.

3. Гибкость (Agility). Способность реагировать на внешние воздействия и изменения рынка и таким образом получать или сохранять конкурентное преимущество. Метрики гибкости включают адаптивность и стоимостную оценку риска.

4. Затраты (Costs). Затраты на обслуживание процессов цепи поставок. Включают затраты на рабочую силу и материалы, управленческие и транспортные расходы. Типичная метрика – себестоимость реализованной продукции.

5. Активы (Assets). Способность эффективно использовать активы. Стратегии управления включают сокращение запасов и использование инсорсинга. Основные метрики – объем запасов в днях и загрузка производственных мощностей.

Первые три группы метрик ориентированы на клиента, последние две – на поставщика.

SCOR-модель предлагает три уровня метрик. Метрики первого уровня диагностируют состояние цепи поставок в целом. Их использование позволяет сформулировать систему целей, которые участники цепи поставок планируют достичь в процессе совместной работы. Метрики второго и третьего уровней дают возможность понять, что определяет эффективность цепи поставок и что на нее влияет.

Моделирование цепочек поставок данных



Рис. 7.3.

Цепочка поставок данных*

* Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Ранее мы уже сравнивали процессы обработки данных с промышленным производством (см. рис. 2.7 и 5.1). Если перенести элементы SCOR-модели на цепочку поставок данных, вот как могут выглядеть основные процессы²⁷⁹:

1. Планировать. Деятельность по планированию должна обеспечивать интеграцию процессов и потоков на протяжении всего жизненного цикла продукта (данных), моделирование сценариев «что, если» для производства, доставки и использования информации, гибкость при нестабильном спросе, понимание вовлеченными лицами и организациями их влияния на последующие и предыдущие элементы цепочки поставки. На этом этапе также планируются спрос, затраты, время выполнения операций, рентабельность данных как актива, методы их учета и многие другие вопросы, связанные с работой с данными.

2. Снабжать. Операции снабжения отвечают за приобретение материалов (данных), за управление запасами (создание и ведение баз данных, работа с метаданными и основными данными [master data]), сетью поставщиков (владельцы данных) и соглашениями с ними (условия предоставления услуг владельцами данных), за обработку платежей и доходов, за передачу и прием данных (пакетная или потоковая передача, включая контроль безопасности), а также за их проверку (профилирование данных,

²⁷⁹ Там же.

контроль качества).

3. Делать. Производственная деятельность обеспечивает превращение материалов (данных) в готовую продукцию (информация в виде отчетов, входные документы для приложений и т. п.). Она включает операции по созданию (очистка, интеграция, обогащение), промежуточную обработку в областях временного хранения, «упаковку» (агрегирование, подготовка отчетов и т. п.) и «выпуск» (загрузка или обновление баз данных).

4. Доставлять. В рамках доставки выполняются операции по управлению заказами (обработка запросов на получение данных и информации), складированию (хранилища данных, витрины данных) и транспортировке (различные виды импорта или экспорта и передачи данных).

5. Возвращать. Деятельность по возврату в информационном контексте может быть не столь актуальна, как в отношении промышленной продукции, поскольку данные возвращаются крайне редко. Однако сюда можно отнести такой важный комплекс мероприятий, как поддержка пользователей данных (в случае ошибок, сбоев и т. п.).

6. Предоставлять возможность. Практически все отмеченные выше обеспечивающие и вспомогательные процессы SCOR-модели в той или иной степени можно рассматривать применительно к цепочке поставок данных. Как мы увидим в части 2, эти процессы реализуются преимущественно в рамках таких функциональных областей управления данными, как руководство данными (data governance), безопасность и качество данных, метаданные.

Для оценки цепочек поставок данных по характеристикам эффективности, определенным в SCOR-модели, можно использовать, например следующие метрики:

1. Надежность. Производительность выполнения запросов, регулярность проведения обновлений, качество данных (точность, полнота, своевременность, целостность и т. п.).

2. Оперативность. Время обработки обращения пользователя, доступность информации, уровень удовлетворенности пользователей.

3. Гибкость. Польза информации для целого ряда задач; показатели использования связанных данных, метаданных и основных данных; простота интеграции новых видов данных.

4. Затраты. Затраты на получение данных, на управление данными, на доставку данных (включая трудозатраты и затраты на технологии).

5. Активы. Актуальность информации, объем доступных исторических данных, фактическое использование (например, процент данных, которые запрашивались пользователями или приложениями).

Приведенное сравнение показывает, что эффективность работы с данными в организации можно существенно повысить за счет внедрения совокупности процессов, аналогичных цепи поставок продукции. При этом многие базовые концепции и подходы допустимо заимствовать у такой хорошо проработанной рамочной структуры, как SCOR-модель, а также у других референтных моделей из области материального производства. Мы продолжим развивать эту тему в части 2, а сейчас в качестве иллюстрации приведем упрощенное представление цепочки поставок данных, предложенное Дагласом Лейни (рис. 7.3).

На рисунке отражены основные элементарные действия, составляющие жизненный цикл данных. Поскольку книга Лейни посвящена вопросам информационной экономики, представленная схема отражает взгляд на цепочку поставок данных исходя из экономических категорий спроса и предложения. Она не содержит такого относительно большого количества групп процессов, как SCOR-модель, и больше напоминает цепочку ценности (см. рис. 7.2). На схеме выделено три основных этапа:

- сбор – каким образом организация получает исходные данные;
- администрирование – какие обеспечивающие повышение ценности данных механизмы может включать информационное производство;
- применение – за счет чего информация приносит выгоду.

Два первых этапа формируют цепочку поставок данных со стороны предложения, третий – со стороны спроса. В ходе сбора и администрирования потенциальная и вероятная ценность данных увеличивается, а на этапе применения она реализуется²⁸⁰.

280 Три уровня ценности информационных активов (реализованная, вероятная и потенциальная) мы рассматривали в главе 5.

Представленные на схеме действия могут быть объединены и упорядочены разными способами. Если одна цепочка поставок пересекается с другой, образуется сеть поставок данных.

Цепочка поставок данных²⁸¹ представляет собой повторяющийся, унифицированный и документированный набор процессов, направленных на повышение ценности данных по мере их прохождения через этапы обработки и передачи лицам, принимающим решения. Реализовать эффективные цепочки стандартных процессов сбора, обработки и предоставления данных могут только грамотные в области данных организации. Чтобы обеспечить надежное и предсказуемое получение высокой отдачи от информационных активов, они должны ориентироваться на внедрение общекорпоративных политик. Процедуры, основанные на привлечении отдельных рабочих групп или проектов, не могут дать нужного результата. Причина проста: у рабочих групп нет стимула совершенствовать методы работы или информационные продукты после того, как они достигнут своих целей. Ценность информационных активов организации устойчиво повышается, только когда разрабатывается и внедряется всеобъемлющая корпоративная логистика цепочки поставок данных. Как только она будет внедрена, организация не должна допускать отклонений от стандартов (за исключением корректировок, проводимых посредством официальных процессов изменений в рамках программы управления организационными изменениями [organizational change management])²⁸².

В этой связи следует рассмотреть такие понятия, как дата-центричная организация и организация, управляемая на основе данных.

7.3. Дата-центричная организация и организация, управляемая на основе данных

Любую цепь поставок продукции можно рассматривать с двух взаимодополняющих друг друга позиций. В центре внимания одной из них активные действия, т. е. совокупность технологических операций, которые составляют процесс. Другая позиция фокусируется на последовательности трансформаций на пути от исходного сырья к готовому изделию; иначе говоря, рассматриваются те преобразования, которые происходят в предмете данной технологии по мере его прохождения по технологической цепочке. Во всех без исключения традиционных технологиях взаимосвязь между этими двумя точками зрения на технологический процесс вполне очевидна и неразрывна прежде всего потому, что предмет обработки материалов, его можно подвергать измерениям, испытаниям и другим необходимым контрольным действиям. К сожалению, в случае цепочки поставок данных сложность состоит в том, что триада «данные, информация и знания», составляющая предмет информационных технологий, нематериальна. Как следствие, процедуры работы с компонентами триады страдают односторонностью. Десятилетиями ИТ были сосредоточены исключительно на разного рода технологических операциях, реализуемых теми или иными устройствами и приложениями. Большинство определений ИТ сводятся к перечислению средств аппаратного и программного обеспечения, служащих для преобразования, хранения, защиты, обработки, передачи и выборки информации, при этом мало говорится о самой информации.

Несмотря на предпринимаемые усилия по развитию прикладных систем для управления бизнесом, подразделения, осуществляющие управление бизнесом, с одной стороны, и ИТ-подразделения предприятий – с другой, остаются разделенными. Причина в том, что различные части информации и данных принадлежат разным владельцам и они по-разному ими оперируют²⁸³.

Хорошо понимая масштабы негативных последствий в сложившейся ситуации, Майкл Петерсон

281 Как уже отмечалось, некоторые авторы, говорят о «цепочке поставок информации». В этой книге мы используем чаще встречающийся термин «цепочка поставок данных».

282 Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

283 Черняк Л. На пути к информационно-центричному предприятию // Открытые системы. СУБД, 2006, 10. – URL: <https://www.osp.ru/os/2006/10/3910094>.

(Michael Peterson), директор по программам развития SNIA²⁸⁴, в 2006 году подготовил меморандум «Информационная конвергенция» (Information Convergence)^{285, 286, 287}.

Петерсон отмечает, что традиционное английское выражение We don't communicate («Мы не общаемся») удачно отражает сложившиеся отношения между лицами, принимающими решения, и представителями ИТ-подразделений. Но под влиянием новых законодательных инициатив, а также в связи с необходимостью учитывать различного рода риски, принимать во внимание требования по повышению оперативности управления и с учетом прочих факторов эта модель дистанцированных отношений, устраивавшая до поры обе стороны, окончательно изжила себя. Новую информационную модель Петерсон назвал информационной конвергенцией, в центре которой находится представление об информационно-центричном предприятии (рис. 7.4).

По определению Петерсона, информационно-центричное предприятие (information-centric enterprise) – это организация, в которой ценность информации и требования к информации используются в качестве основы для управления, администрирования и операций по обеспечению безопасности, что позволяет налаживать тесное сотрудничество между пользователями информации, владельцами информации, операторами и администраторами информационных систем²⁸⁸.

Смысл такого подхода к предприятию в том, что требуется баланс; внимание должно быть сконцентрировано не только на приложениях, интеграции приложений и тому подобных вещах, но и в равной степени на корпоративной информации и данных. Иначе говоря, нужно совместить операционный и предметный взгляд на природу вещей, как это делается в более традиционных технологиях.

284 SNIA – Ассоциация индустрии сетей хранения (Storage Networking Industry Association) – некоммерческая организация, объединяющая ведущих производителей продуктов для сетей хранения данных (в их числе Quantum, Storage Tek, IBM, Intel, Seagate и др.), <https://www.snia.org/>.

285 Там же.

286 Long D., Reis L. Information Convergence: At the Boundaries of Access: Introduction, 25 J. Marshall J. Computer & Info. L. 585. 2008. – URL: <https://repository.law.uic.edu/jitpl/vol25/iss4/1/>.

287 Peterson M. Collaborate or Die! Data Management Forum: White Paper. Storage Networking Industry Association, 2007. – URL: https://www.snia.org/sites/default/files/SNIA-DMF_Collaborate-or-Die_20070120.pdf.

288 Там же.



Рис. 7.4.

Информационно-центричное предприятие*

* Черняк Л. На пути к информационно-центричному предприятию // Открытые системы. СУБД, 2006, 10. – URL: <https://www.osp.ru/os/2006/10/3910094>.

В основу информационно-центричного предприятия должна быть положена единая информационная инфраструктура, которая позволит принимать принципиально важные решения на уровне предприятия в целом, преодолевая внутриведомственные барьеры. Эта инфраструктура может стать общим знаменателем для конвергенции управления информацией с информационными технологиями и средствами информационной безопасности, позволяя каждому из направлений оставаться ответственным за себя, но работать в едином бизнес-контексте.

Но не технологии главное: более существенно налаживание организационного взаимодействия между такими участниками производственного процесса, как руководители бизнес-подразделений, финансовые работники, юристы, менеджеры, отвечающие за информацию, сотрудники ИТ-подразделений и специалисты по информационной безопасности. Цель этого объединения – классификация информации и данных, а также определение требований к работе с ними²⁸⁹.

В последнее время гораздо чаще, чем термин «информационно-центричное предприятие», встречается термин «дата-центричная организация» (data-centric organization). Кроме того, пожалуй, еще чаще говорят об организации, управляемой на основе данных (data-driven organization).

Если первые два термина применительно к большинству контекстов вполне взаимозаменяемы, то

²⁸⁹ Черняк Л. На пути к информационно-центричному предприятию // Открытые системы. СУБД, 2006, 10. – URL: <https://www.osp.ru/os/2006/10/3910094>.

между терминами «дата-центричная организация» и «организация, управляемая на основе данных» существуют различия. Рассмотрим их более подробно. Начнем с обсуждения особенностей организации, управляемой на основе данных.

Карл Андерсон (Carl Anderson) в книге «Создание организации, управляемой на основе данных»²⁹⁰ пишет, что «управление на основе данных подразумевает формирование инструментов, способностей и, что самое важное, корпоративной культуры, которая опирается на данные». Организация должна создать у себя так называемую аналитическую цепочку ценностей, т. е. отработать последовательность прохождения этапов от сбора данных до принятия решений²⁹¹.

Хотя задача сбора различных видов данных из разных источников для принятия более эффективных решений очень важна, но организация не всегда находит способ эти данные стандартизировать и синхронизировать. В результате она получает скопление разрозненных, трудноконтролируемых сведений. Некоторые организации накапливают огромные информационные массивы, для манипулирования которыми и выявления полезных трендов требуются специальные серверы или обработка данных в облаке — это может стать серьезно мешающим фактором.

²⁹⁰ На эту книгу часто ссылаются при обсуждении вопросов управления на основе данных. В русском переводе она называется «Аналитическая культура. От сбора данных до бизнес-результатов»*.

* Anderson C. Creating a Data-Driven Organization. O'Reilly Media, 2017. (Русский перевод: Андерсон К. Аналитическая культура. От сбора данных до бизнес-результатов. — М.: Манн, Иванов и Фербер, 2017.)

²⁹¹ Anderson C. Creating a Data-Driven Organization. O'Reilly Media, 2017. (Русский перевод: Андерсон К. Аналитическая культура. От сбора данных до бизнес-результатов. — М.: Манн, Иванов и Фербер, 2017.)

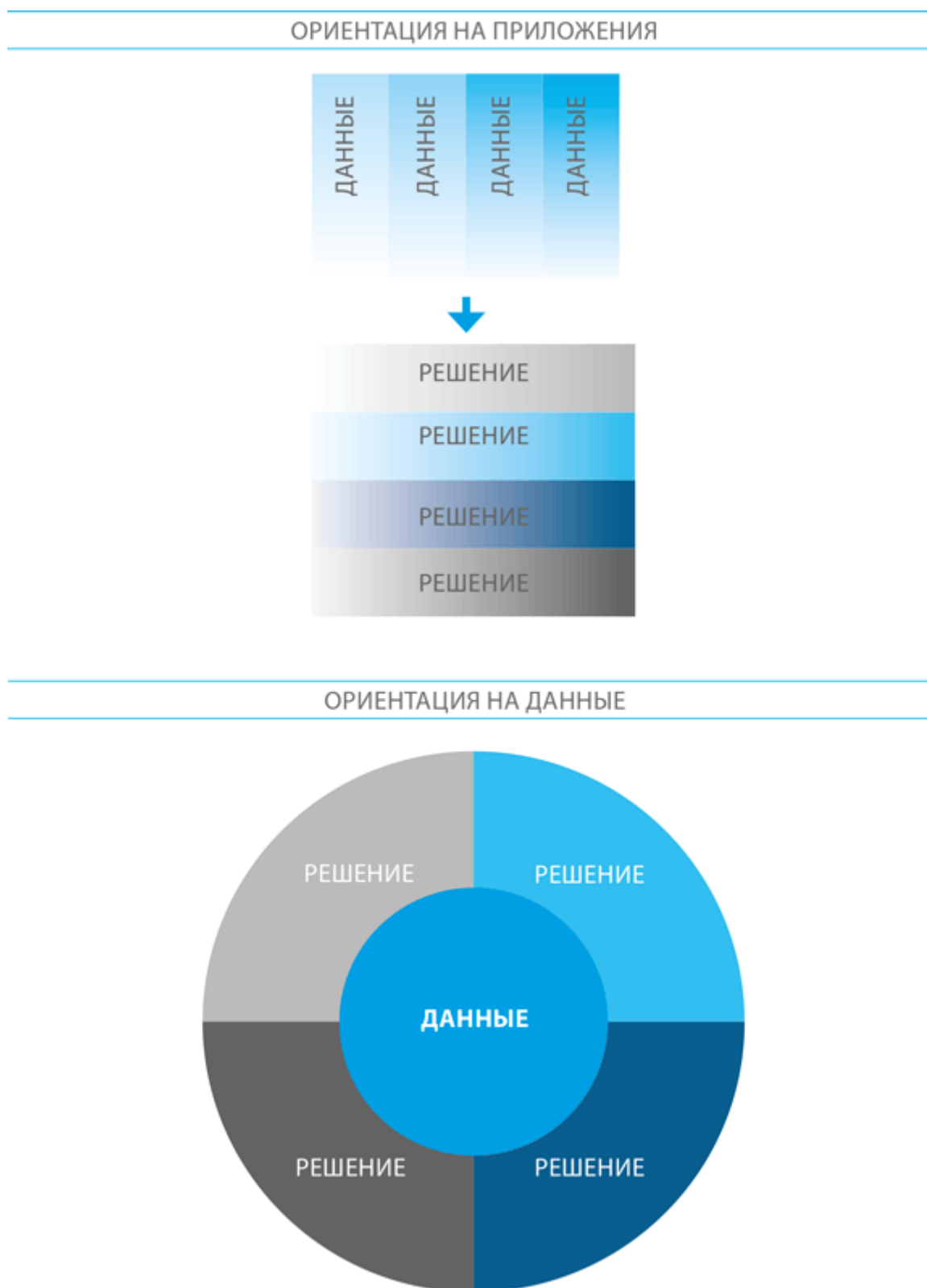


Рис. 7.5.

Ориентация архитектуры: на приложения и на данные*

* Dunn C. The Difference Between Data-centric and Data-driven. Applied Software Technology Inc., 2021. – URL: <https://www.asti.com/the-difference-between-data-centric-and-data-driven/>.

К сожалению, в большинстве организаций наиболее распространена практика, когда программные приложения, применяемые для анализа при принятии решений, размещают обрабатываемые данные в

своих отдельных хранилищах – «бункерах» (silo)²⁹². При этом другие приложения не могут их использовать. Вместо того чтобы ориентироваться на данные, архитектура информационных систем ориентирована на приложения²⁹³. Даже если процесс принятия решения в такой организации и будет обеспечивать управление на основе данных, она не будет дата-центричной. Может даже возникнуть ситуация, когда с возрастанием степени управляемости на основе данных степень дата-центричности будет уменьшаться^{294, 295, 296}.

Чтобы стать дата-центричной, организации необходимо разработать единую корпоративную модель данных, а затем привести данные в соответствие с этой моделью. Программные приложения, которые анализируют и обрабатывают данные, могут с течением времени меняться, но данные остаются постоянным организационным активом (рис. 7.4)^{297, 298}.

В 2015 году Дэйв Маккомб²⁹⁹ и команда руководимого им агентства Semantic Arts, более 20 лет курирующего построение ИТ-архитектур для крупных организаций, опубликовали Манифест дата-центричности (The Data-Centric Manifesto). Предназначение манифеста – дать огласку проблеме неэффективного распоряжения данными (Маккомб говорит о «дата-центричной революции») и наметить путь к ее решению^{300, 301}.

В манифесте заявляется о зависимости многих организаций от приложений. Годами компании хранили информацию в базах данных бизнес-приложений. В итоге их подразделения имеют разрозненные массивы информации, а реализация интеграционных решений требует огромных усилий и финансовых вложений.

Поставив в центр системы корпоративные данные, компании перестанут быть зависимыми от прикладного ПО и его разработчиков. Появится возможность включать в систему предприятия любые новые приложения и исключать старые.

В манифесте нет конкретных решений по преобразованию ИТ-ландшафта. Каждая организация выбирает свой путь перехода к дата-центричной архитектуре. Тем не менее в манифесте содержатся

²⁹² Понятие silo мы уже рассматривали в предыдущей главе.

²⁹³ Этот вопрос мы уже затрагивали в главе 6, когда говорили о дата-центричном мышлении и рассматривали Доктрину в области данных.

²⁹⁴ Doubleday K. Introduction to Data-Centricity. Fluree PBC | Medium, 2020. – URL: <https://medium.com/fluree/introduction-to-data-centricity-61d73286b7f2>.

²⁹⁵ McComb D. The Data-Centric Revolution: Data-Centric vs. Data-Driven. The Data Administration Newsletter (TDAN.COM), 2016. – URL: <https://tdan.com/the-data-centric-revolution-data-centric-vs-data-driven/20288>.

²⁹⁶ Shyamalkumar D., Subramanya J. Data-Centric Organization – A New Paradigm. Whitepaper. Larsen & Toubro Infotech Limited (LTI), 2021. – URL: <https://www.lintinfotech.com/wp-content/uploads/2020/07/Data-Centric-Organization-A-New-Paradigm.pdf>.

²⁹⁷ Dunn C. The Difference Between Data-centric and Data-driven. Applied Software Technology Inc., 2021. – URL: <https://www.asti.com/the-difference-between-data-centric-and-data-driven/>.

²⁹⁸ Vista Projects. Data-centric Architecture – A Different Way of Thinking. Vista Projects Ltd., 2021. – URL: <https://www.vistaprojects.com/blog/data-centric-architecture/>.

²⁹⁹ Дэйв Маккомб (Dave McComb) – один из главных популяризаторов концепции дата-центричной архитектуры *.

* ТриниДата. Дата-центричная архитектура. ТриниДата, 2021. – URL: https://trinidata.ru/data_centric_architecture.htm.

³⁰⁰ McComb D. The Data-Centric Revolution. The Data Administration Newsletter (TDAN.COM), 2015. – URL: <https://tdan.com/the-data-centric-revolution/18780>.

³⁰¹ Semantic Arts. The Data-Centric Manifesto. Semantic Arts Inc., 2015. – URL: <http://datacentricmanifesto.org/>.

принципы, которые позволят встать на путь извлечения выгоды из корпоративных данных³⁰².

Из Манифеста дата-центричности

Принципы

- Данные являются важнейшим активом любой организации.
- Большинство современных систем ориентировано на приложения.
- Хранение данных в проприетарном ПО – ошибка.
- Дороговизна и сложность корпоративных систем связана с отношением приложений к данным.
- Мы понимаем, что ориентация на приложения приносит деньги...
- ...но дата-центричный подход принесет больше.

Способы думать о дата-центричной революции:

Сейчас: ориентация на приложения	В будущем: ориентация на данные (дата-центричность)
Непомерно высокие затраты на изменения в бизнес-приложениях	Разумная стоимость изменений
Данные привязаны к приложениям, потому что приложения владеют данными	Данные — открытый ресурс, который переживает любое приложение
Каждый новый проект сопровождается большими усилиями по преобразованию данных	Каждый новый проект использует существующие хранилища данных
Данные существуют в широком разнообразии разнородных форматов, структур, значений и терминологии	Данные глобально интегрированы, имеют общее значение и экспортируются из общего источника в любой необходимый формат
Интеграция данных потребляет 35–65% ИТ-бюджета	Интеграция данных всегда будет бесплатной
Трудно или невозможно интегрировать внешние данные с внутренними данными	Внутренние и внешние данные легко интегрируются

7.4. Управление информационными активами и управление корпоративной информацией: в чем разница?

Приведенные выше рассуждения показывают, что в основу управления информационными активами организации должны быть положены принципы дата-центричности и управления на основе данных. Только их соблюдение позволит извлечь из данных максимальную ценность. Проясним несколько ключевых моментов, связанных с терминологией, касающейся подходов к такому управлению.

В литературе последних лет наиболее часто можно встретить следующие термины:

- управление данными (Data Management, DM);
- управление информацией (Information Management, IM);
- управление корпоративной информацией (Enterprise Information Management, EIM);
- управление информационными активами (Information Asset Management, IAM).

Как эти термины соотносятся друг с другом?

302 ТриниДата. Дата-центричная архитектура. ТриниДата, 2021. – URL: https://trinidata.ru/data_centric_architecture.htm.

Согласно определению DAMA-DMBOK³⁰³, управление данными – это разработка, выполнение и контроль выполнения политик, программ и практик предоставления, проверки, защиты и повышения ценности данных и информационных активов на протяжении всего их жизненного цикла.

Поскольку в DMBOK термины «информация» и «данные» используются (с известными оговорками) как взаимозаменяемые синонимы³⁰⁴, можно сказать, что термины «управление данными» и «управление информацией» также являются синонимами.

Перейдем к термину «управление корпоративной информацией» (EIM).

Как отмечает Джон Лэдли, определение DMBOK для DM является общим и может распространяться на деятельность по управлению данными как на локальном уровне (отдельные подразделения организации), так и на корпоративном (организация в целом). Когда говорится об управлении корпоративной информацией, имеется в виду именно управление на уровне всей организации³⁰⁵.

Важно еще раз обратить внимание на то, что и управление информацией (IM и DM) и управление корпоративной информацией (EIM) обычно рассматриваются как программы, т. е. совокупности взаимосвязанных проектов и другой деятельности, направленных на достижение общей цели и реализуемых в условиях общих ограничений³⁰⁶.

Разберемся теперь с термином «управление информационными активами».

IAM – постоянно уточняющийся и расширяющийся набор принципов и концептуальных положений, определяющих подход к управлению данными как стратегически важному активу организации³⁰⁷. Основные аспекты IAM достаточно полно отражены в книге Дагласа Лейни «Инфономика: информация как актив: монетизация, оценка, управление». В частности, там говорится, что IAM можно рассматривать как новый, дата-центричный образ мышления. Чтобы стать дата-центричной и управляемой на основе данных, организация должна «разбираться в вопросах информации» (такие организации называют *information-savvy*), в частности хорошо понимать концепции IAM³⁰⁸.

Разница во взглядах на данные как ресурс и актив еще раз укрупненно представлена на рисунке 7.6.

Важная особенность понятия «управление корпоративной информацией» – то, что оно рассматривается в неразрывной связи с понятием IAM. IAM предоставляет концепции, а EIM – это

303 Далее в этой главе и во второй части книги мы будем по возможности использовать в качестве базиса определения из DAMA-DMBOK. Такого же подхода при обсуждении вопросов управления данными предлагает придерживаться ряд известных специалистов, в частности Джон Лэдли в своей книге «Руководство данными: Как разработать, внедрить и поддерживать эффективную программу руководства данными»*.

* Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

304 Это обсуждалось в конце главы 2.

305 Там же.

306 См. обсуждение «смертного греха» № 3 в главе 6.

307 В качестве примеров элементов IAM можно привести предложенные Джоном Лэдли «Общепринятые принципы учета информационных активов» (см. главу 5), а также Лидерский манифест о данных (см. главу 6).

308 Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

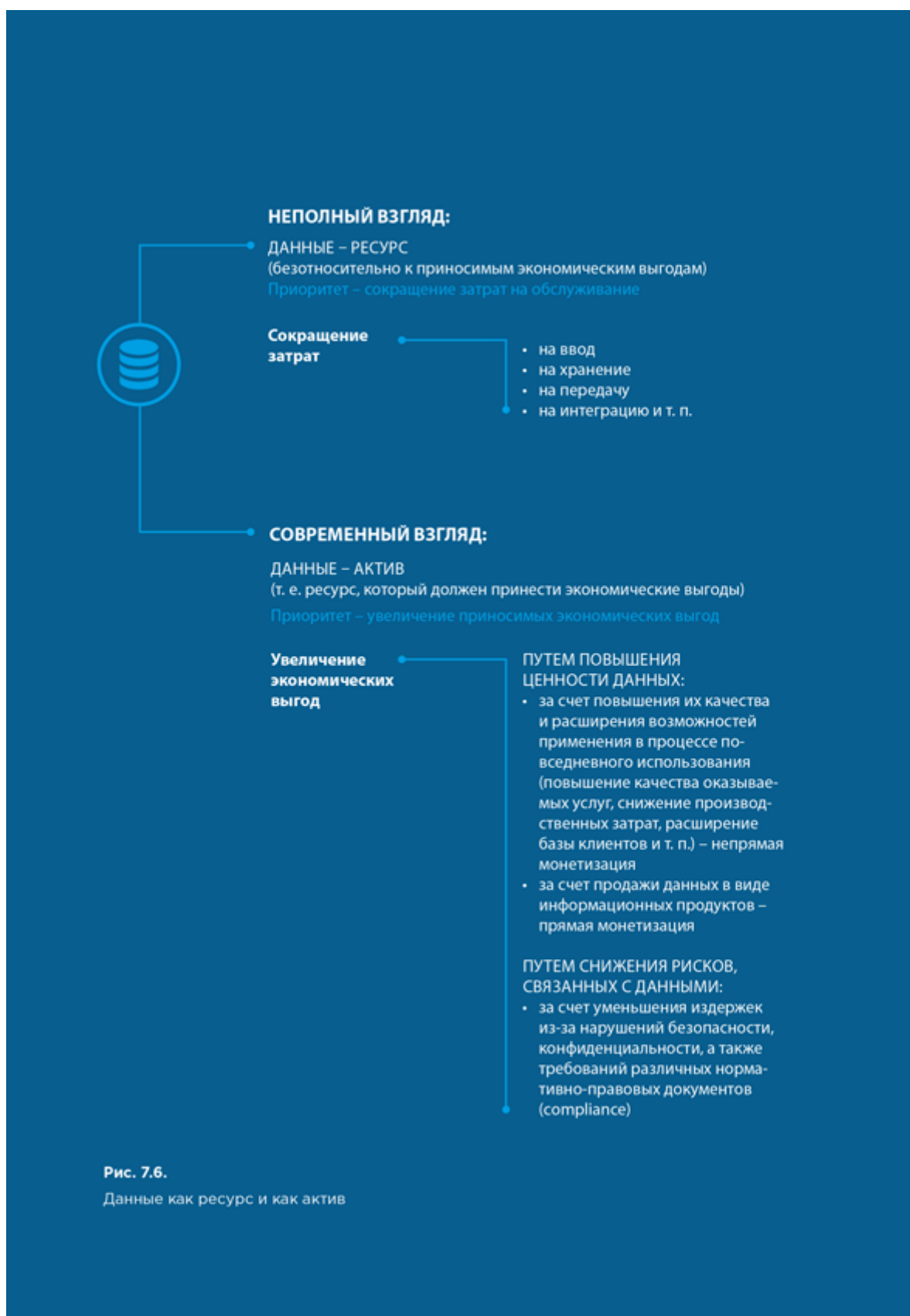
программа, реализующая эти концепции^{309, 310, 311}.

Выше мы уже говорили о том, что реализовать эффективные цепочки поставок данных в рамках отдельных рабочих групп или проектов невозможно. Необходимо организовать всеобъемлющую корпоративную информационную логику, т. е. внедрить программу ЕАМ. Следует заметить, что, хотя положения ДМВОК носят универсальный характер и могут быть применены в организации как на локальном, так и на корпоративном уровне, все же основной акцент в них сделан на уровне корпорации. Далее в этой книге управление данными мы будем обсуждать именно в контексте управления корпоративной информацией.

³⁰⁹ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

³¹⁰ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

³¹¹ Evans N., Price J. Enterprise information asset management: the roles and responsibilities of executive boards // Knowledge Management Research & Practice, 14: 353–361. – URL: <https://www.experiencematters.com.au/wp-content/uploads/2018/03/Enterprise-information-asset-management-the-roles-and-responsibilities-of-executive-boards.pdf>.



7.5. Управление корпоративной информацией: необходимость актуализации повестки дня для совета директоров

В главе 5 мы говорили о документе «Информация как актив: повестка дня для совета директоров», который был выпущен в 1995 году так называемым Комитетом Хоули под эгидой компании KPMG. Многие

организации, сотрудничающие с KPMG, внедрили у себя рекомендации из предложенной Комитетом Хоули повестки. Однако со временем интерес к ней ослаб. В 2017 году KPMG совместно с британским Привилегированным институтом библиотечных и информационных работников (Chartered Institute of Library and Information Professionals, CILIP)³¹² и рядом других заинтересованных организаций запустили программу подготовки обновленной версии документа, подготовленного Комитетом Хоули. По мнению участников программы, данный документ абсолютно не утратил своей актуальности. Если на первый поставленный Комитетом Хоули вопрос: «Располагает ли ваша организация информацией, которая являлась стратегическим активом?» – большинство менеджеров высшего и среднего звена сегодня отвечает: «Да», то на второй вопрос: «Понимают ли они эти активы и управляют ли ими так же, как понимают другие свои стратегические активы, используя их и защищая должным образом?» – многие по-прежнему отвечают: «Нет».

С момента выхода повестки Комитета Хоули появился ряд новых факторов, которые требуют своего отражения и в то же время делают ее еще более актуальной.

Необходимость цифровой трансформации побуждает организации анализировать данные для извлечения ценности, реализации инноваций и поиска новых возможностей. В последние годы происходит заметное перемещение взгляда на информацию от традиционных сценариев использования в рамках области ИТ к рассмотрению ее в более широкой роли ключевого аспекта выполнения бизнес-стратегии^{313, 314}.

В 2019 году был выпущен документ «Информация как актив: Сегодняшняя повестка дня для совета директоров». По структуре он схож с повесткой Комитета Хоули – содержит перечень рекомендаций, а также чек-лист и пояснительные примечания, описывающие мероприятия, необходимые для их выполнения. Кроме того, в документе описан ряд определяющих факторов и принципов³¹⁵.

«Сегодняшняя повестка дня для совета директоров» подтверждает принцип, согласно которому лидерство и модели поведения в области управления корпоративной информацией (EIM) должны исходить от высшего руководства организации.

Во введении мы уже говорили о том, что четвертая промышленная революция, основанная на цифровизации, характеризуется дизруптивным (ломающим привычные представления) воздействием на утвердившиеся традиционные компании. По сравнению с контекстом предыдущей повестки сегодня совет директоров и менеджмент организации должны учитывать ряд новых факторов³¹⁶.

• **Искусственный интеллект, машинное обучение и автоматизация.** Развитие искусственного интеллекта трансформирует способность информационных систем извлекать выгоду из информационных активов. Благодаря машинному обучению, анализу данных и роботизации информационные системы могут обучаться быстрее людей и предлагать для организации аналитику и новые идеи, намного превосходящие по результативности прежние возможности.

• **Рост объема и охвата.** В последнее десятилетие наблюдается беспрецедентный рост объема и спектра информационных активов, больших данных, полученных в результате увеличения числа подключенных устройств. Организации инвестируют средства в системы, которые могут хранить

³¹² <https://www.cilip.org.uk/>.

³¹³ CILIP/KPMG. Information as an Asset: Today's Board Agenda. CILIP/KPMG Position Paper. February 2019. London: CILIP. – URL: <https://www.cilip.org.uk/page/informationasset>.

³¹⁴ Ward S., Carter D. Information as an asset – Today's Board Agenda: The value of rediscovering gold // Business Information Review, 2019, 36 (2): 53–59. DOI: 10.1177/0266382119844639 – URL: https://www.researchgate.net/publication/332763457_Information_as_an_asset_-_Today's_Board_Agenda_The_value_of_rediscovering_gold.

³¹⁵ CILIP/KPMG. Information as an Asset: Today's Board Agenda. CILIP/KPMG Position Paper. February 2019. London: CILIP. – URL: <https://www.cilip.org.uk/page/informationasset>.

³¹⁶ Там же.

разнообразные наборы данных, предоставлять точные метаданные и объединять внутренние данные с внешними, полученными извне организации, для решения задач анализа.

● **Сетевое взаимодействие.** Современные подходы к сетевому обмену данными позволяют организациям находить и использовать информационные ресурсы как находящиеся внутри, так и предоставляемые внешними источниками, обмениваться ими, что приводит к более глубокому пониманию решения бизнес-задач и новым возможностям.

● **Защита информационных активов во взаимосвязанном мире.** В нашем все более взаимосвязанном мире безопасность и сохранность информационных активов часто зависит от поставщиков услуг за пределами организации, таких как поставщики облачных систем хранения и обработки данных. Это затрудняет непосредственный контроль защиты информационных активов со стороны совета директоров. Надлежащий надзор за управлением данными (data governance) – ключевая область, требующая пристального внимания.

● **Конфиденциальность данных.** Широкие обсуждения скандалов, связанных с утечками персональных данных в сети, подчеркивают растущую осведомленность общественности и политиков об информационных рисках. Недостатки корпоративного управления персональными данными вызывают справедливую критику и могут оказать влияние на бизнес-результаты.

Потенциал и риски современной информационной среды должны быть хорошо сбалансированы, что тем самым будет стимулировать инновации. Возрастающая сложность управления операционными рисками, обеспечения нормативно-правового соответствия и безопасности данных, повышения эффективности бизнес-процессов и качества обслуживания клиентов требует от совета директоров лидерства и выработки общего видения.

7.6. Управление корпоративной информацией: сегодняшняя повестка дня для совета директоров

Приведем пункты «Сегодняшней повестки дня для совета директоров», сформулированные с учетом отмеченных факторов^{317, 318}.

Совет директоров должен убедиться в том, что его собственная деятельность осуществляется таким образом, что члены совета:

- уверены в том, что информация, которую они используют для разработки стратегии и принятия решений, достаточна и необходима для их целей;
- имеют возможность получать доступ и запрашивать всю информацию и подтверждающие данные, необходимые в качестве основы для их решений, и уверены, что эта информация предоставляется объективно;
- уверены в том, что практика использования ими информации, как на коллективном, так и на индивидуальном уровне соответствует применимым законам, нормативным актам и признанным этическим стандартам (национальным и международным);
- решительно настроены поддерживать осведомленность о новых разработках в области управления информацией, ресурсах и инструментах. Этого можно достичь, возложив ответственность за управление информацией на конкретного директора или старшего менеджера, который отчитывается об управлении информацией непосредственно на уровне совета.

Если приведенные выше принципы соблюдаются в работе совета директоров, то это естественным образом обуславливает его ответственность за стратегию и политику в области работы с информацией.

1. Совету нужно выработать информационно-центричное видение бизнеса. Для реализации такого видения ему необходимо обеспечить разработку и регулярный пересмотр стратегии организации в области

317 Там же.

318 Ward S., Carter D. Information as an asset – Today's Board Agenda: The value of rediscovering gold // Business Information Review, 2019, 36 (2): 53–59. DOI: 10.1177/0266382119844639 – URL: https://www.researchgate.net/publication/332763457_Information_as_an_asset_-_Today's_Board_Agenda_The_value_of_rediscovering_gold.

работы с информацией (information strategy). Стратегия должна быть направлена на увеличение ценности, которую можно извлечь из корпоративных информационных активов и направить на поддержание текущей и будущей бизнес-стратегии. Она откроет для организации новые возможности, основанные на использовании данных, и будет способствовать капитализации разработок в информационной среде. Сегодня компаниям необходимо быть среди лидеров. Анализ данных должен предупреждать о конкурентных угрозах и выявлять потенциальные направления роста и инновационной деятельности.

2. Совету необходимо рассмотреть и принять политику в области работы с информацией (information policy), охватывающую все аспекты управления ею, критически важные для деятельности организации.

– *Идентификация.* Организация должна уметь идентифицировать информационные активы и проводить различие между теми, что представляют ценность и важность, и теми, что этого не делают. Совету директоров следует отдавать себе отчет в том, что ценность информации может меняться по мере того, как меняется бизнес-среда.

– *Ценность.* Политика должна способствовать выявлению возможностей получения организацией выгоды (за счет правовой защиты, лицензирования, повторного использования, объединения, публикации или приобретения информационных активов).

– *Руководство данными* (data governance)³¹⁹. Политика работы с информацией должна включать частные политики (и предусматривать внедрение соответствующих процедур), направленные на надлежащее (в соответствии с применимыми правовыми, нормативными, операционными и этическими нормами и стандартами) использование информации. Сюда же относится и контроль за соблюдением требований по защите информации (включая информацию, принадлежащую другим лицам) от кражи, потери, несанкционированного доступа, злоупотреблений и неправильного использования.

– *Качество и полнота.* Политика должна определять, каким образом организация будет обеспечивать уровни качества и полноты данных, необходимых ей для достижения своих целей на каждом уровне ответственности. В частности, целесообразно предусмотреть введение корпоративных стандартов данных, способствующих эффективной передаче знаний и информации. Кроме того, следует уделять внимание выработке корпоративной информационной архитектуры, которая обеспечит доступность и интеграцию данных.

– *Лучшие практики.* Политику необходимо направить на применение передовых практик в области управления данными, документами и записями. Этот аспект включает определение ролей и обязанностей по созданию, сбору, организации, хранению, предоставлению доступа, использованию и защите информации. Нужно предусмотреть комплекс мер по обучению сотрудников соответствующим навыкам, в том числе и подходам к применению информационных активов. В рассматриваемых практиках так или иначе следует задействовать каждого сотрудника.

– *Культура.* Политика должна способствовать формированию организационной культуры, которая станет стимулировать сотрудников шире использовать имеющиеся информационные активы (а также вводить в действие новые) и выявлять возможности, основанные на интеграции информации и беспрепятственном обмене идеями³²⁰.

3. Большую часть повседневных обязанностей по управлению информацией нужно делегировать сотрудникам по всей организации, при этом их выполнение следует подкрепить четкой операционной структурой (operational framework) зон ответственности руководителей, команд и отдельных работников, прозрачной на всех уровнях и утвержденной советом директоров. Она будет залогом эффективного руководства данными и выполнения бизнес-планов (дорожных карт), определяющих инициативы в области управления информацией.

4. Совет директоров должен быть уверен, что информационная политика соответствует целям организации, т. е. иметь возможность определять, каким образом будет оцениваться ее соблюдение.

На рисунке 7.7 схематически отражена рамочная модель управления корпоративной информацией. Она демонстрирует, каким образом основные описанные элементы сочетаются между собой.

³¹⁹ Деятельность в области data governance подробно рассматривается в главе «Руководство данными» (глава 10).

³²⁰ О важности изменения организационной культуры мы уже говорили в главе 6, когда рассматривали «смертный грех» № 7.

Бизнес-стратегия определяет видение и стратегию организации в области работы с информацией. Выделенные в виде опорных колонн базовые компоненты согласованно взаимодействуют, обеспечивая реализацию видения и стратегии. Эффективность реализации оценивается с помощью метрик. Частные политики, процессы, процедуры и стандарты, поддерживающие политику работы с информацией, опираются на информационную архитектуру, формируемую в соответствии с дата-центричным подходом. Вместе все эти составляющие содействуют развитию информационно-центричной культуры и обуславливают существенное повышение конкурентоспособности организации за счет эффективного использования ее информационных активов.



Рис. 7.7.

Рамочная модель управления корпоративной информацией*

* Ward S., Carter D. Information as an asset – Today’s Board Agenda: The value of rediscovering gold // Business Information Review, 2019, 36 (2): 53–59. DOI: 10.1177/0266382119844639 – URL: https://www.researchgate.net/publication/332763457_Information_as_an_asset_-_Today's_Board_Agenda_The_value_of_rediscovering_gold.

7.7. Управление корпоративной информацией: выгоды от внедрения и чек-лист для совета директоров

Авторы «Сегодняшней повестки дня для совета директоров» выделяют следующие основные выгоды, получаемые организацией от внедрения программы управления корпоративной информацией³²¹.

1. Повышение эффективности и производительности операционной деятельности (нужная

³²¹ CILIP/KPMG. Information as an Asset: Today’s Board Agenda. CILIP/KPMG Position Paper. February 2019. London: CILIP. – URL: <https://www.cilip.org.uk/page/informationasset>.

информация в нужное время):

- улучшенное планирование;
- более обоснованное принятие решений;
- повышенная гибкость;
- улучшенные результаты;
- расширение сотрудничества;
- улучшенное сетевое взаимодействие;
- более мотивированный персонал.

2. Повышение отдачи от инвестиций организации в технологии:

- снижение затрат;
- более целенаправленные инвестиции;
- гарантированный долгосрочный доступ к информации;
- сокращение времени и усилий.

3. Повышение эффективности использования информационных активов в целом по организации:

- более быстрые и точные ответы на запросы;
- расширение корпоративных знаний;
- более интенсивное использование активов;
- расширение возможностей применения;
- создание прочной основы для сотрудничества.

4. Повышение эффективности бизнеса:

- создание конкурентных преимуществ;
- расширение возможностей для роста и диверсификации;
- более качественный и эффективный сервис;
- упрощение бизнес-процессов;
- ускорение принятия решений;
- обмен лучшими практиками;
- сокращение дублирования;
- повышение рентабельности инвестиций (ROI).

5. Повышение креативности и расширение инновационной деятельности:

- новые идеи;
- развитие творческой культуры;
- благоприятная инновационная среда.

6. Ускорение реагирования и повышение конкурентоспособности:

- чувствительность к изменениям;
- поддержка развития бизнеса;
- опора на бизнес-анализ.

7. Обеспечение нормативно-правового соответствия (compliance):

- снижение рисков;
- улучшение подотчетности;
- улучшение соблюдения требований законодательных и других нормативных документов;
- сохранение корпоративной памяти;
- повышение открытости и прозрачности.

Несмотря на перечисленные выгоды, убедить топ-менеджеров в необходимости изменения отношения к данным и переходу к управлению ими как корпоративным активом бывает непросто. Поэтому к «Сегодняшней повестке дня...» (как и к повестке, выпущенной Комитетом Хоули в 1995 году) приложен контрольный перечень (чек-лист). Он представляет собой набор вопросов, которые призваны заострить внимание высшего руководства организации на ключевых аспектах управления корпоративной информацией и оценить текущее положение. Вопросы разбиты на три группы.

1. Уверены ли вы и ваш совет директоров в том, что ваша организация никогда не столкнется со следующими проблемами:

- риски крупных штрафов и корпоративных неприятностей из-за неправильного использования

информации;

- разочарование ваших сотрудников и возникновение помех в их работе из-за плохого информационного обеспечения (низкие надежность и качество данных, отсутствие стандартов данных, препятствия в доступе к информации);

- нереализованные возможности повышения эффективности за счет совершенствования методов управления информацией;

- отставание от конкурентов, которые управляют информацией и используют ее более успешно, чем вы;

- упущенные возможности развития вашего бизнеса в новых направлениях, появляющиеся благодаря использованию информации?

2. Уверены ли вы и ваш совет директоров в том, что:

- вас всегда точно и полно информируют о связанных с данными проблемах и рисках в области инвестиционных решений и бизнес-планов;

- вы имеете доступ к объективной, точной и актуальной информации, на основе которой можно принимать эффективные решения;

- вы всегда используете возможности для реализации инноваций, направленных на развитие информационного обеспечения;

- вы регулярно извлекаете выгоду из использования информационных активов, с помощью которых добавляется ценность для клиентов, рынков, сотрудников и партнеров;

- вы знаете обо всех местах хранения данных и информации вашей организации?

3. Полностью ли вы и ваш совет директоров удовлетворены состоянием дел в части следующих элементов информационного обеспечения:

- у организации есть стратегия работы с информацией, поддерживающая бизнес-стратегию;

- в организации разработаны и действуют политики, процессы и практики, направленные на реализацию этой стратегии;

- бизнес-процессы организации, включая работу с персоналом и развитие культуры, соответствуют стратегии работы с информацией;

- результаты работ по вышеуказанным направлениям регулярно проверяются и пересматриваются на уровне совета директоров?

Как видим, совет директоров призван сыграть центральную роль в обеспечении управления корпоративными информационными активами таким образом, чтобы они создавали ценность, стимулировали инновации и открывали новые возможности.

Во второй части книги мы более подробно поговорим о современных подходах к формированию в организации полнофункциональной системы управления корпоративной информацией, позволяющей выстраивать эффективные цепочки ценности и поставок данных.

Литература к главе 7

- Мерзляк А. В. Роль информации и стратегии в моделях управления цепями поставок: референтные модели лучших практик, Ментцера, GSCF, CPFR, SCOR. Российское предпринимательство, 2015. 16(22), 4099–4118. DOI: 10.18334/rp.16.22.21.

- DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

- Davenport T., Evgeniou T., Redman T. Your Data Supply Chains Are Probably a Mess. Here's How to Fix Them // Harvard Business Review. 2021. – URL: <https://hbr.org/2021/06/datamanagement-is-a-supply-chain-problem>.

- Evans N., Price J. Barriers to the Effective Deployment of Information Assets: An Executive Management Perspective // Interdisciplinary Journal of Information, 7, 2012: 177–199. DOI: 10.28945/1721 – URL: https://www.researchgate.net/publication/289400090_Barriers_to_the_Effective_Deployment_of_Information_Assets_An_Executive_Management_Perspective.

- Kotorov R. The Data Value Chain: Steps for Monetizing Your Data. Integration Developer News, 2017. – URL: <https://www.idevnews.com/stories/6998/The-Data-Value-Chain-Stepsfor-Monetizing-Your-Data>.
- Ladley J., McGilvray D., O’Neal K., Price J., Redman T. The Leader’s Data Manifesto. dataleaders.org, 2019. – URL: <https://dataleaders.org/>.
- Legner C., Pentek T., Otto B. Accumulating Design Knowledge with Reference Models: Insights from 12 Years’ Research into Data Management // Journal of the Association for Information Systems, 2020, 21(3): 735–770. DOI: 10.17705/1jais.00618 – URL: https://www.researchgate.net/publication/341684789_Accumulating_Design_Knowledge_with_Reference_Models_Insights_from_12_Years%27_Research_into_Data_Management.
- The Hawley Committee. Information as an Asset: The Board Agenda. London: KPMG, 1995. – URL: <https://www.cilip.org.uk/page/informationasset>.
- Vista Projects. Data-centric Architecture – A Different Way of Thinking. Vista Projects Ltd., 2021. – URL: <https://www.vistaprojects.com/blog/data-centric-architecture/>.

Часть 2. Данные: Извлечение ценности От планирования к расширению возможностей применения

Глава 8. Данные как объект управления

8.1. Источники данных и виды информационных активов

Организации, которые не знают, какими данными они располагают, не могут использовать их в качестве актива. В книге Дагласа Лейни «Инфономика: информация как актив: монетизация, оценка, управление» приводится справедливое высказывание директора по информационным технологиям крупной страховой компании: «Глупо, что у кого-то в компании есть опись нашей офисной мебели, но ни у кого нет описи того, какими данными мы располагаем»³²².

При инвентаризации информационных активов целесообразно в первую очередь разделить их на группы в зависимости от источников поступления данных. Лейни выделяет пять основных групп (рис. 8.1).

Операционные данные

Это данные о клиентах, поставщиках, партнерах и сотрудниках, доступные в процессе онлайн-обработки транзакций и (или) полученные из онлайн базы данных аналитической обработки. Часто такие сведения успешно собираются с помощью датчиков в ходе мониторинга процессов предприятий. Например, кассовые аппараты, подключенные к банковской системе, интеллектуальные счетчики, голосовая связь, радиочастотная идентификация и т. д.

«Темные (dark) данные»

Информация, которая не хранится или не собирается организациями специально, а формируется случайно в процессе ведения бизнеса или взаимодействия с сетевыми сервисами и остается в интернет-архивах. Такие данные являются общедоступными и частично структурированными для анализа, включают электронные письма, электронные договоры, документы, мультимедиа, системные журналы и т. д.

³²² Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

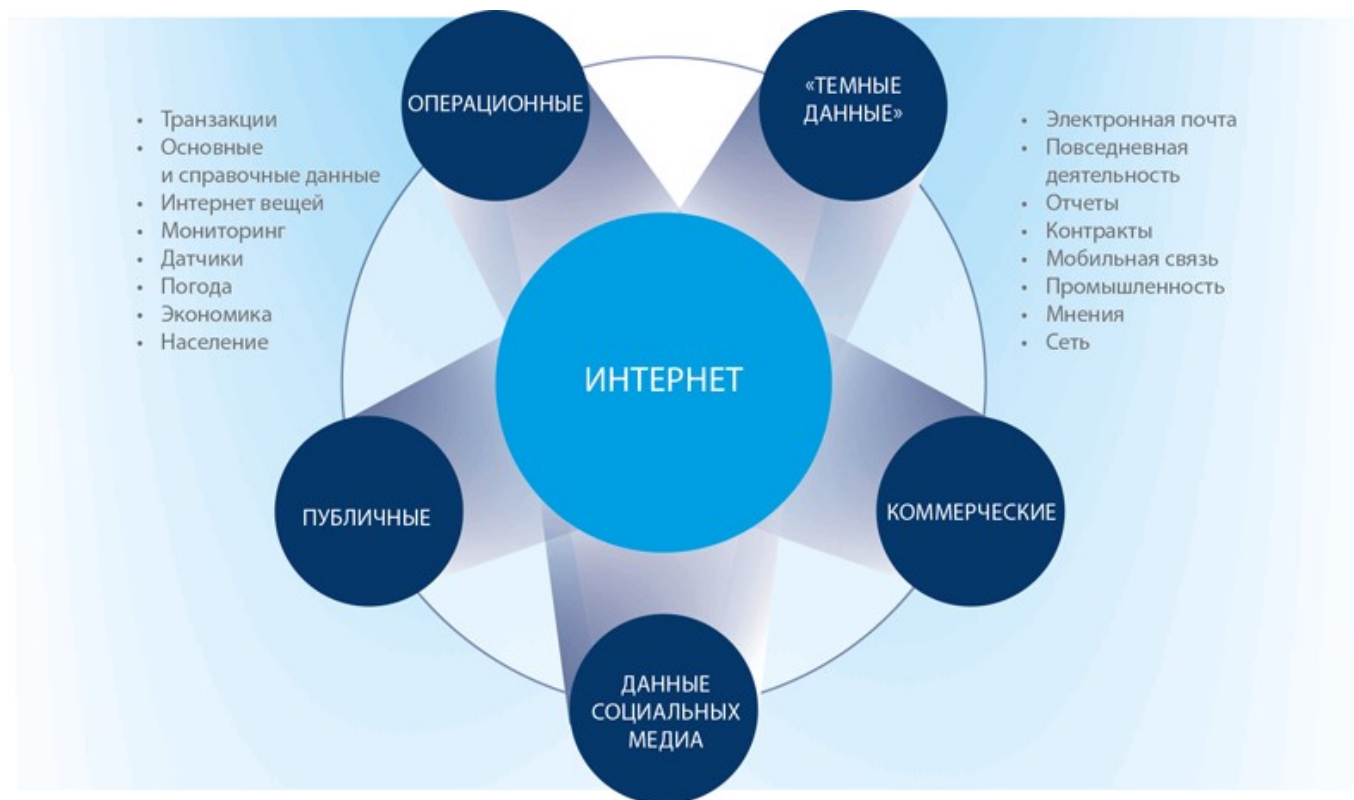


Рис. 8.1.

Информационные активы организации*

* Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Даглас Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Публичные данные

Информация, распространяемая государственными органами (заявления, пресс-релизы, прогноз погоды, сведения о планах муниципального развития; открытые публичные реестры, опубликованные нормативные акты, включая их проекты), одна из наиболее достоверных и чаще всего структурированная. Ценность таких данных раскрывается в совокупности с другими источниками сведений, поскольку позволяет определить направления развития бизнеса или целой индустрии в рамках отдельного города, страны или на международном уровне.

Коммерческие данные

Уже давно в разных отраслях промышленности существуют агрегаторы коммерчески ценной информации. Указанные агрегаторы предоставляют полный доступ к собственным каталогам информации по подписке. Но с учетом перенаправления современных рыночных отношений в сторону открытия информации для потенциальных инвесторов и клиентов многие сведения, представляющие коммерческий интерес, открыто размещаются в цифровой среде. Распространенной стала практика размещения информации об активах на открытых площадках, в особенности если речь идет о принадлежащих компаниям объектах интеллектуальной собственности.

Данные социальных медиа

Вовлеченность бизнеса и частных лиц в функционал крупных социальных сетей создала еще один источник данных о спросе, тенденциях в определенных сегментах рыночных отношений, новых и перспективных продуктах, услугах и компаниях. Сообщения, комментарии, репосты активно используют для выявления и прогнозирования целевых клиентов, коммерческих возможностей, конкурентных отношений, бизнес-рисков и потенциальных партнеров.

Открытые данные

Эта категория данных на рисунке 8.1 не отражена, поскольку она тесно связана с категорией публичных данных. Термин «открытые данные» появился в 1995 году в американском научном сообществе в виде призыва свободно обмениваться данными. Несмотря на общую открытость публичных и открытых данных, между ними существует принципиальная разница. Она заключается в том, что использование публичных данных определяется законом – доступ к ним можно получить, например, по специальному запросу. Суть открытых данных в обратном – данные должны быть опубликованы еще до того, как кому-то понадобятся^{323, 324}.

8.2. Классификация данных

На практике при организации управления данными их обычно классифицируют по следующим признакам.

По назначению и области применения обычно выделяют:

- **метаданные** – данные, описывающие структуру и характеристики данных;
- **справочные данные** – данные из справочников, международных, общероссийских и отраслевых классификаторов и т. п.;
- **основные данные** – структурированные данные об объектах учета;
- **транзакционные данные** – сведения, отражающие результат изменения данных, относящиеся к фиксированному моменту времени, не изменяющиеся в будущем;
- **данные контроля и аудита** – сведения, фиксируемые в различных журналах регистрации^{325, 326, 327}.

Часто в отдельную категорию относят **аналитические данные** – эти данные фактически образуются из основных, справочных и транзакционных данных. Они используются в аналитической деятельности организации (рис. 8.2).

На рисунке 8.2 отражены взаимоотношения перечисленных категорий данных в процессе деятельности организации.

³²³ Измestьева Е. История открытых данных. Теплица социальных технологий, 2013. – URL: <https://te-st.ru/2013/04/09/brief-history-open-data/>.

³²⁴ Chignard S. A brief history of Open Data. ParisTech Review, 2013. – URL: <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>.

³²⁵ McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM). Morgan Kaufmann, 2008.

³²⁶ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

³²⁷ Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

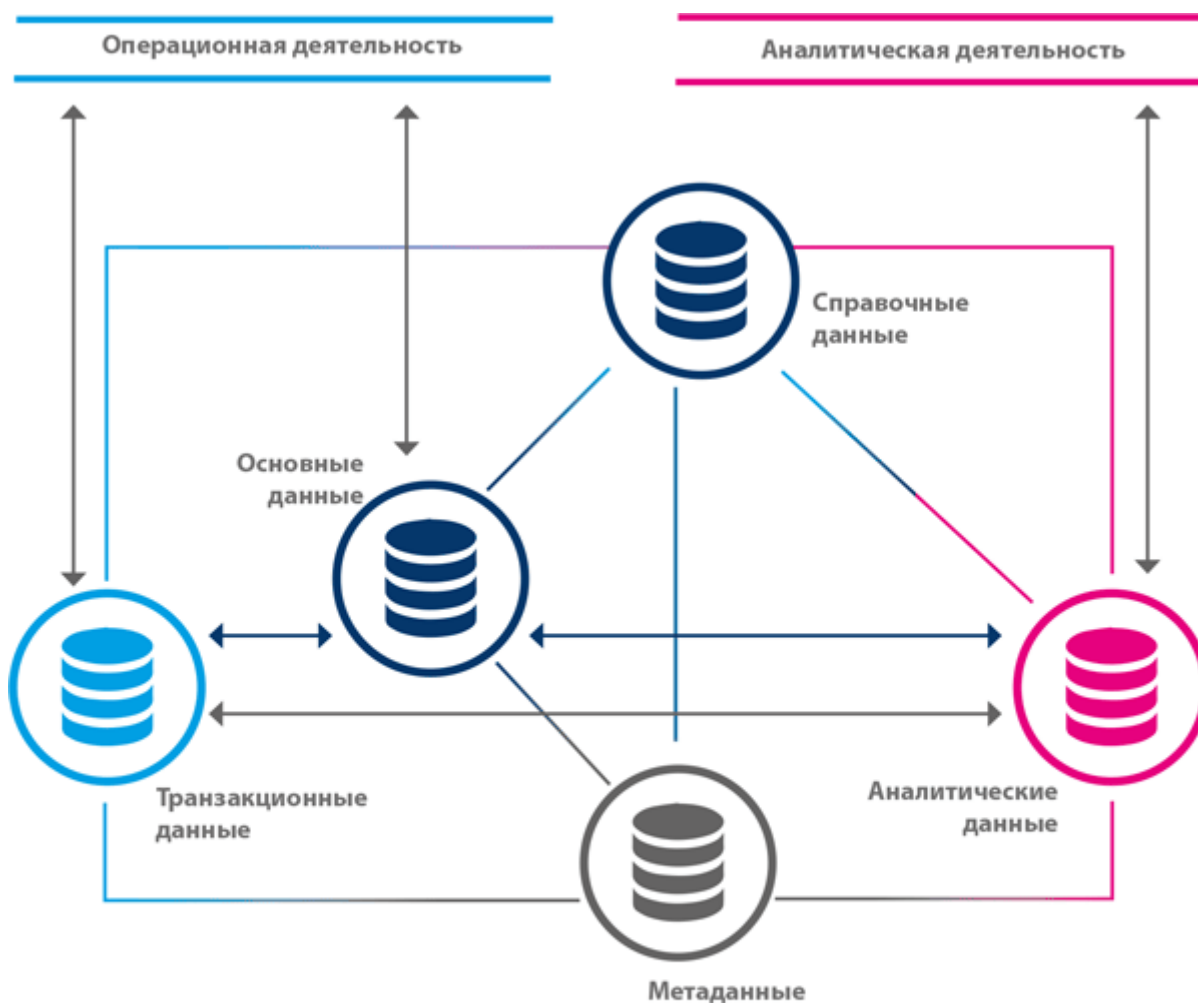


Рис. 8.2.

Взаимоотношения основных категорий данных в процессе деятельности организации*

* Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

На рисунке 8.3 отражены роли, которые играет каждая из категорий данных в информационном обеспечении процессов организации. Следует обратить внимание на фундаментальную роль справочных и основных данных и на важность поддержания высокого уровня их качества. Например, при наличии ошибок в данных о номере товара или типе клиента цена заказа на доставку может быть определена некорректно (см. связи, отраженные пунктирными стрелками), что может привести к серьезным финансовым последствиям.

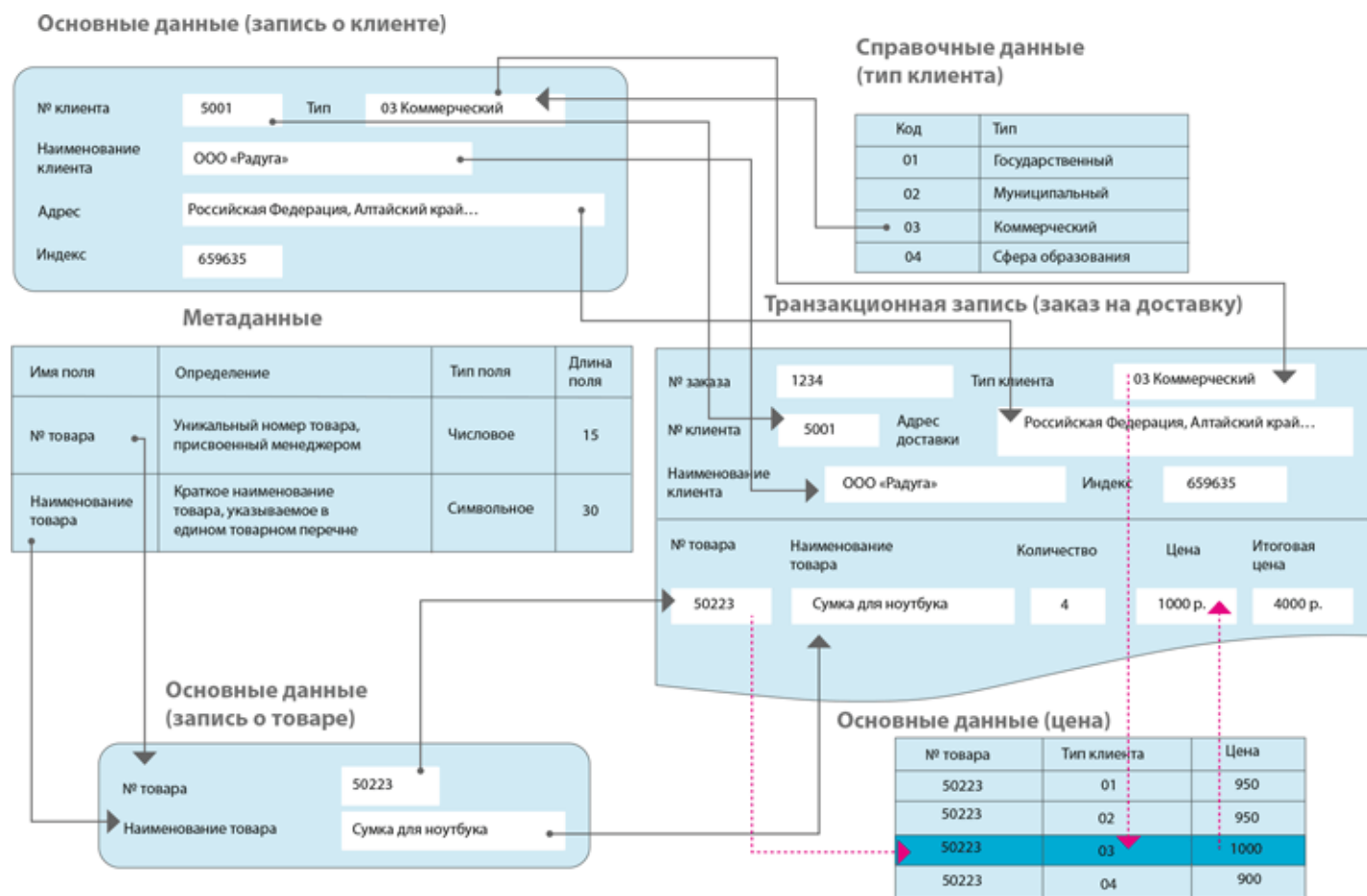


Рис. 8.3.

Роли отдельных категорий данных в информационном обеспечении процессов организации (на основе материалов*)

* McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM). Morgan Kaufmann, 2008.

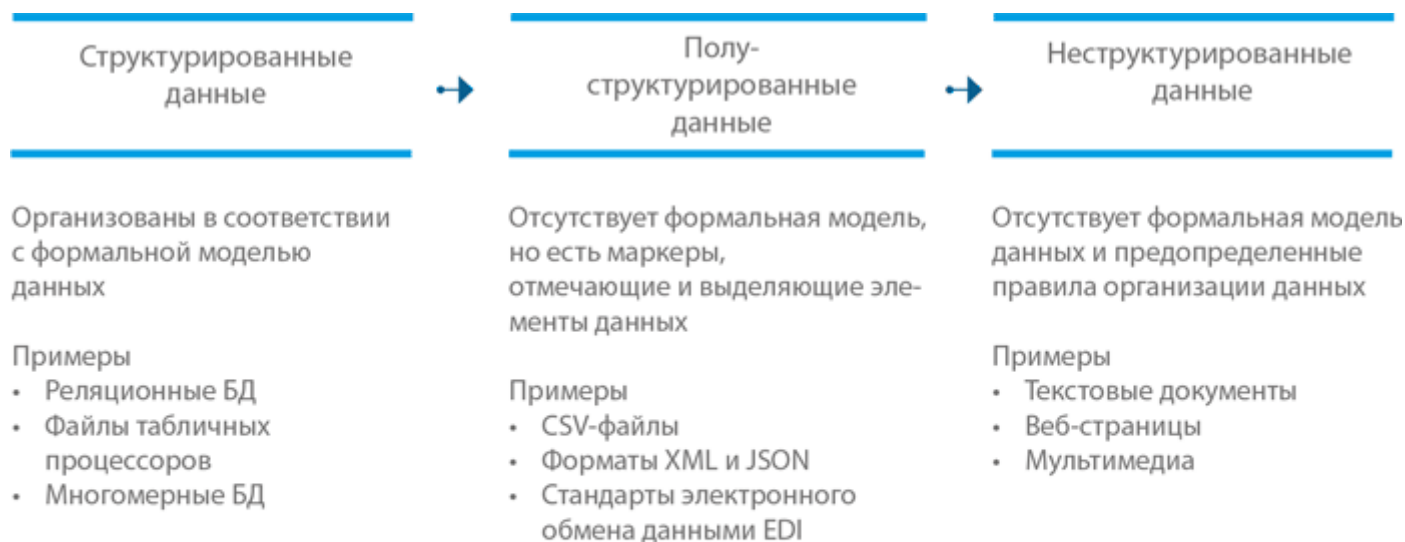


Рис. 8.4.

Форматы хранения и передачи данных с разной степенью структурированности
(на основе материалов*)

* Deng Z. MIS2502: Data Analytics: Semi-structured Data Analytics. Fox School of Business. Temple University, 2019. – URL: <https://slidetodoc.com/mis-2502-data-analytics-semistructured-data-analytics-zhe/>.

По степени структурированности можно выделить:

• **структурированные данные** – данные, имеющие строго фиксированную структуру, определяемую формальной моделью данных (например, реляционной схемой³²⁸);

• **полуструктурированные данные** – данные, не имеющие строго определенной структуры, но предполагающие наличие установленных правил, позволяющих выделять семантические элементы при их интерпретации (прежде всего, правил расстановки тегов и других маркеров, отмечающих и выделяющих элементы данных);

• **неструктурированные данные** – данные, произвольные по форме, не имеющие строго определенной структуры и не организованные по определенным правилам.

Схемы, представленные на рисунках 8.2 и 8.3, в основном отражают взаимосвязи между структурированными данными. Однако в деятельности предприятий и учреждений не менее важны данные полуструктурированные и неструктурированные (в частности, к ним относятся отмеченные выше данные контроля и аудита). Они могут быть самыми разнообразными по назначению и области применения. С каждым годом роль этих данных становится все более заметной и существенной.

На рисунке 8.4 приведены примеры форматов хранения и передачи данных по каждой из перечисленных категорий.

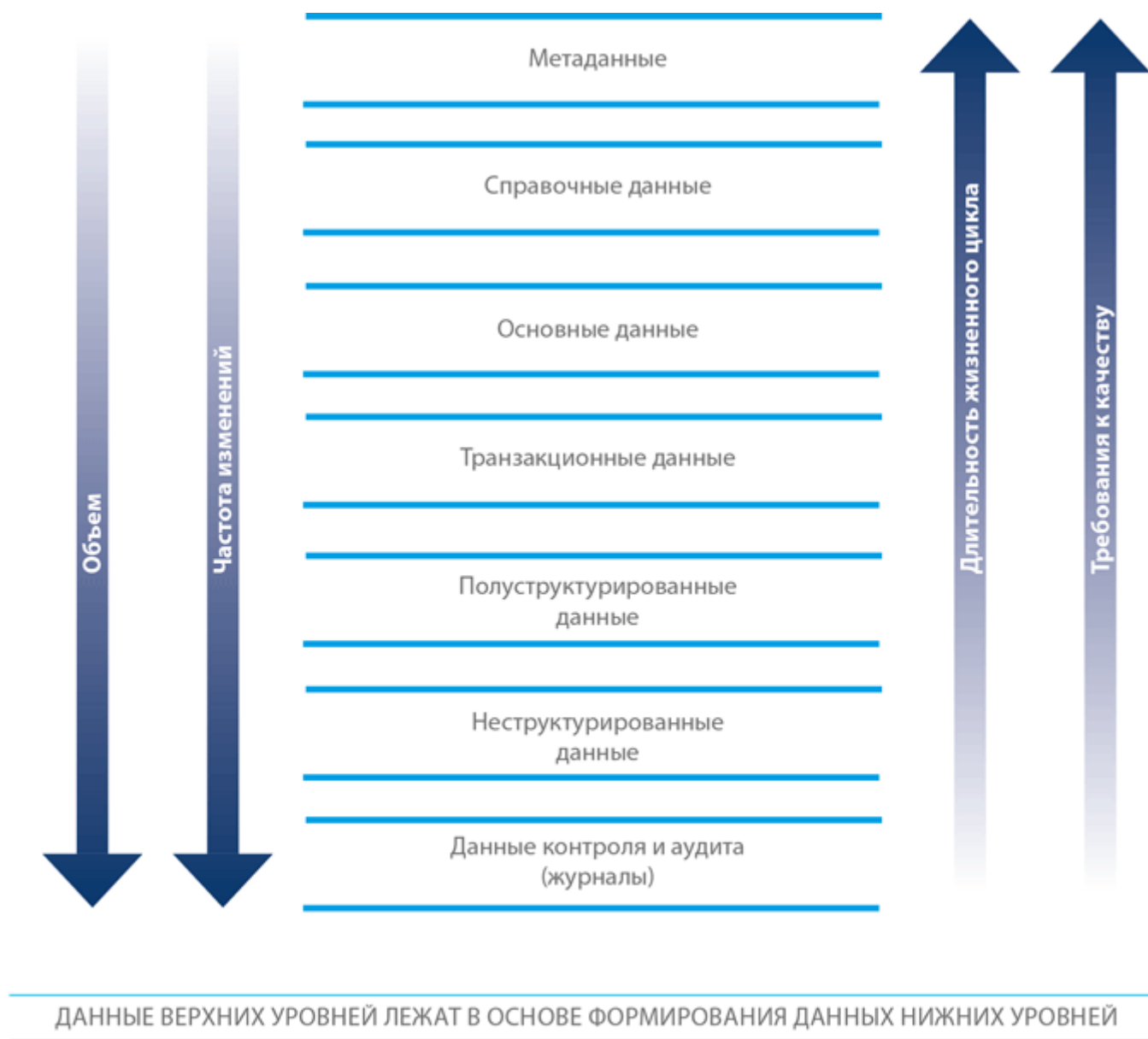


Рис. 8.5.

Соотношения между категориями данных*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

С точки зрения управления данными полезно представить их в виде диаграммы (рис. 8.5), укрупненно отражающей соотношения между основными категориями³²⁹.

Данные, относящиеся к категориям, расположенным сверху, как правило, являются базовыми для формирования данных, относящихся к категориям, расположенным ниже (данные верхних категорий участвуют в формировании данных нижних категорий). Поэтому по мере продвижения вверх по списку категорий требования к качеству соответствующих данных возрастают.

Также по мере продвижения вверх по списку категорий увеличивается продолжительность жизненного цикла данных. При этом при продвижении вниз по списку категорий увеличивается объем

³²⁹ На диаграмме объединены категории, выделяемые по разным признакам (назначению и степени структурированности), однако их можно упорядочить, поскольку метаданные, справочные, основные и транзакционные данные являются преимущественно структурированными.

самих данных, а также частота их изменений.

Говоря о данных контроля и аудита, следует отдельно определить такие категории данных, как машинные данные и потоковые данные.

К машинным данным относится информация, автоматически генерируемая компьютером, процессом, приложением или устройством без вмешательства человека. Они становятся одним из основных источников информации, а это в первую очередь относится к данным контроля и аудита.

Потоковые данные могут относиться почти к любой из перечисленных выше категорий, однако у них имеется одно дополнительное свойство. Данные поступают в систему непрерывно по мере возникновения тех или иных событий, а не загружаются в хранилище данных дискретно большими массивами.

К особой категории можно отнести большие данные (big data). Термин «большие данные» связан преимущественно с техническими аспектами формирования и обработки. Он не предполагает конкретные виды данных (эта категория может включать и структурированные, и неструктурированные, и полуструктурированные данные). Традиционно принято определять большие данные по трем признакам (3V): Volume, Velocity, Variety^{330, 331}. Коротко о них скажем.

● **Volume – объем.** К 2020 году общий объем информации, созданный в цифровой среде, достиг 44 ЗБ. По прогнозам Всемирного экономического форума, к 2025 году объем ежедневного интернет-трафика данных по всему миру достигнет 463 ЭБ. С точки зрения наглядной оценки такого огромного объема информации следует отметить, что для его записи потребуется больше 212 млн DVD-дисков. Информация, которая образует объем больших данных, поступает от миллионов используемых электронных сетевых устройств и приложений. Важно иметь в виду, что на этапе накопления big data отбора ненужных данных не производится. Обычные инструменты хранения и анализа не способны справиться с таким объемом.

● **Velocity – скорость.** Указанные выше объемы данных поступают в обработку в режиме реального времени, в отличие от традиционной обработки пакетов данных. Это означает, что они накапливаются моментально, при этом не имеет значения продолжительность потока самих данных. Таким образом, при работе с большими данными не только фиксируются их потоки, но и производится их запись и обработка в таком виде, чтобы не было потерь.

● **Variety – разнообразие.** Большие данные формируются из поступающих от различных источников сведений в разнообразных форматах (видеоданные, фотографии, звуковые записи, текстовые сообщения, файлы транзакций, комментарии, использование ссылок и фиксация просмотров страниц и т. д.). Наибольший объем составляют полуструктурированные и неструктурированные данные социальных сетей и социальных медиасервисов. Таким образом, термин big data не относится исключительно к большим данным в понимании объема. Он значительно шире, поскольку подразумевает также большие скорости поступления данных и большое разнообразие источников и форматов получаемой информации.

Со временем правило 3V в отношении больших данных стали расширять за счет дополнительных признаков^{332, 333, 334}, в частности:

● **Veracity – достоверность.** Из-за большого объема и вариативности источников поступающих данных сложно проконтролировать их достоверность. Соответствие, точность и правдивость получаемой информации могут быть подтверждены только в результате тщательного анализа и сопоставления.

● **Variability – вариативность.** При обработке и сопоставлении исходные значения полученных

330 Ульянова Е. В. Big Data: технология, принципы и архитектура // Журнал Суда по интеллектуальным правам, 2020, 4 (30): 32–41. – URL: <http://ipcmagazine.ru/legal-issues/big-data-technology-principles-and-architecture#11>.

331 Firican G. The 10 Vs of Big Data. TDWI Upside, 2017. – URL: <https://tdwi.org/pages/upside.aspx>.

332 Следует заметить, что дополнительные признаки (V), добавляемые к первым трем, как правило, описывают уже не характеристики больших данных, а проблемы, возникающие в связи с их использованием.

333 Ульянова Е. В. Big Data: технология, принципы и архитектура // Журнал Суда по интеллектуальным правам, 2020, 4 (30): 32–41. – URL: <http://ipcmagazine.ru/legal-issues/big-data-technology-principles-and-architecture#11>.

334 Firican G. The 10 Vs of Big Data. TDWI Upside, 2017. – URL: <https://tdwi.org/pages/upside.aspx>.

данных могут меняться. В первую очередь данный признак проявляется при работе с речевыми и текстовыми данными. Для понимания точного значения отдельных слов необходима разработка сложных программных продуктов, позволяющих определять смысловую нагрузку исходя не только из прямого значения, но и из контекста.

● **Visualization – визуализация.** Полученные в результате сбора данные непригодны для восприятия человеком. Поэтому требуется их обработка для представления в доступной форме – визуализация. Характерный пример визуализации данных – построение графиков и диаграмм, отображающих результаты анализа данных. Важна возможность самостоятельной настройки. Необходимые параметры представления пользователи определяют сами, в зависимости от поставленных целей и задач.

● **Value – ценность.** Потенциальная ценность больших данных крайне высока. На ценность влияют тщательный и точный анализ данных, актуальность информации и полученные в результате визуализации выводы. Наибольший коммерческий и научный интерес представляют те сведения, которые можно использовать для решения текущих задач конкретного пользователя, а также результаты анализа, которые способствуют генерации новых идей.

Наконец, в зависимости от носителя данных, могут быть выделены:

- данные на бумажных носителях;
- данные в электронном виде.

8.3. Жизненный цикл данных, цепочка данных и происхождение данных

Как и у любого другого актива, у данных есть свой жизненный цикл. Для эффективного управления информационными активами организации необходимо его понимание и планирование.

Концептуально жизненный цикл данных описывается достаточно просто (рис. 8.6). Он включает:

- процессы планирования, а также проектирования и обеспечения доступности данных;
- процессы, которые создают или получают данные;
- процессы, которые осуществляют их перемещение, преобразование, хранение, а также обеспечивают обслуживание данных и предоставление совместного доступа к ним;
- процессы использования и расширения возможностей применения данных;
- процессы, обеспечивающие их ликвидацию.

Кроме того, на протяжении всего их жизненного цикла данные могут очищаться, преобразовываться, подвергаться слиянию или агрегироваться.

Требования к организации отдельных этапов жизненного цикла могут существенно различаться в зависимости от вида данных. Поскольку к данным разных категорий предъявляются различные требования, им присущи различные риски и отведены различные роли в организации, многие инструменты управления данными всецело сфокусированы на различных аспектах классификации и контроля. Например, основные данные имеют иное назначение и области применения, нежели транзакционные, соответственно и требования к управлению данными двух этих категорий предъявляются различные.

Специфика конкретного жизненного цикла данных в отдельно взятой организации может оказаться весьма запутанной, поскольку в течение цикла данные обычно перемещаются из одного места в другое внутри организации, а также за ее пределами. По сравнению с остальными видами активов ситуация с данными усложняется за счет такой характеристики, как возможность совместного использования неограниченным количеством потребителей³³⁵. Поэтому при управлении данными, наряду с управлением на отдельных фазах их жизненного цикла, важно обеспечивать контроль их движения по различным участкам хранения и обработки.

335 См. первый закон информации, который обсуждался в главе 5.



Рис. 8.6.

Жизненный цикл данных*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Для обозначения набора процессов и систем, участвующих в перемещении данных, часто используются термины «цепочка данных» (data chain) и «информационная цепочка» (information chain)³³⁶. Хотя у них есть не менее часто употребляемые синонимы – «поток данных» и «информационный поток»³³⁷, метафора цепочки может оказаться очень полезной. Она подразумевает связь процессов и систем. Если одно звено обрывается, это негативно сказывается на всей последовательности элементов. При этом заметим, что для отражения реальных взаимосвязей в большинстве организаций (когда один набор данных может быть звеном многих цепочек) более точной была бы метафора сети данных.

С жизненным циклом данных и цепочкой данных связано такое понятие, как «происхождение данных» (data lineage). Под происхождением (lineage) обычно понимается линия связи с предком. Большинство людей, интересующихся происхождением данных, хотят осознать два аспекта.

Во-первых, их интересует подтвержденная информация о самом раннем экземпляре (первоисточнике) данных. В искусстве для обозначения истории владения художественным произведением (с того момента, как оно было создано, и по настоящее время) используется термин «провенанс» (provenance). В отношении данных существует аналогичный термин data provenance. Его можно перевести просто как «провенанс данных».

Во-вторых, люди хотят знать, как (а иногда и почему) данные менялись в процессе перехода от самого раннего экземпляра. Изменения могут происходить внутри одной системы или при передаче между системами. Понимание изменений в данных требует понимания цепочки данных, правил, которые применялись к данным по мере их перемещения по цепочке, и того, какое влияние эти правила оказали на данные.

Происхождение данных (data lineage) включает в себя и их провенанс, и сведения об изменениях данных (сведения о последовательности шагов по изменению при движении по цепочке данных, в том числе и при подготовке к применению для различных целей). Используя метафору цепочки легко представить, что данные по мере своего перемещения по ее звеньям будут сохранять некоторые (но не все) свойства своих предыдущих состояний и приобретать новые в процессе преобразования.

³³⁶ Цепочку данных не следует путать с цепочкой поставок данных. В отличие от цепочки поставок это понятие используется только для описания движения данных и не охватывает процессы, связанные с повышением их ценности.

³³⁷ Потоки данных рассматриваются в главах «Планирование и проектирование данных» (глава 11) и «Обеспечение доступности и обслуживание данных: основы» (глава 12).

Следует заметить, что трактовки понятий data chain, data lineage и data provenance в разных источниках могут несколько различаться. Часто data lineage и data provenance рассматриваются отдельно. Здесь мы ориентируемся на книгу выпускающего редактора DMBOK2 Лауры Себастьян-Коулман «Измерение качества данных в целях постоянного совершенствования: Рамочная модель для оценки качества данных»³³⁸.

Наличие подробной информации о происхождении дает возможность проводить анализ влияния на данные (data impact analysis) – выяснять, какие элементы данных в целевой базе данных или в приложении будут затронуты, если мы изменим тот или иной элемент в предшествующих звеньях цепочки данных. На рисунке 8.7 представлен простейший пример описания происхождения элемента данных. Мы видим, что элемент «Сумма заказа», физически реализованный в базе данных как столбец `zz_total`, зависит от трех связанных с ним элементов: «Цена за шт.» (`yy_unit_cost`), «Скидка» (`yy_disc`) и «Заказано (шт.)» (`yy_qty`).



Рис. 8.7.

Пример схематического представления происхождения элемента данных*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

8.4. Системы и люди (организации, подразделения, сотрудники) в процессах управления данными

Как мы уже отмечали, наличие у данных такой отличительной характеристики, как возможность совместного использования неограниченным количеством потребителей, обуславливает их частое перемещение между информационными системами. При рассмотрении процессов управления данными целесообразно разделить системы на несколько укрупненных категорий в соответствии с их ролями в информационных цепочках³³⁹.

Обратимся еще раз к рисунку 8.2. Он поможет составить представление об основных видах информационных систем, которые обычно функционируют в организациях.

Прежде всего нужно выделить целевые системы (target systems). Наиболее часто к таким системам относятся базы и хранилища аналитических данных, используемые в аналитической деятельности организации.

Целевые системы получают данные из исходных систем (source systems). Преимущественно в этом качестве выступают операционные (транзакционные) системы, обеспечивающие деловые операции

³³⁸ Sebastian-Coleman L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework (The Morgan Kaufmann Series on Business Intelligence). Morgan Kaufmann, 2013.

³³⁹ Там же.

(транзакции). Они поддерживают такие функции, как создание записей для клиентов, обработка заказов и оплата счетов. Как правило, цель этих систем – не хранение исторических данных для использования в анализе и при принятии решений (хотя иногда бывает и так). Их назначение – обеспечение возможности совершения бизнес-транзакций.

Термины «целевая система» и «исходная система» относительны. Целевая система, данные из которой непосредственно используются в рамках одной функциональной области, может быть исходной системой для хранилища данных, используемого в другой области.

Обрабатываемые сведения обычно поступают по сложной цепочке данных, поэтому между исходными системами целесообразно проводить границу. Например, система, в которой данные были первоначально созданы, выделяется как система-первоисточник или система происхождения (system of origin), а система, из которой данные фактически поступают в целевое хранилище, называется системой – непосредственным источником (direct source system).

Отдельно следует выделить системы записи (systems of record) – системы, которые отвечают за хранение наиболее полных и достоверных описаний объектов того или иного вида. В практике управления основными данными такие описания называются «золотыми записями» (golden records), а системы записи также называются «системами истины» (system of truth)³⁴⁰.

В главе 7 мы говорили об управлении данными в контексте формирования цепочек поставок данных. Теперь кратко рассмотрим основные укрупненные роли, которые могут играть как системы, так и люди (подразделения организаций или отдельные сотрудники, а также организации в целом) именно с точки зрения цепочек поставок. Тут нужно обратить внимание на следующие позиции³⁴¹.

Производители данных (data producers)

Производители данных – это люди и системы, которые данные создают. Данные могут создаваться специально с целью предоставления для использования или генерироваться как побочный продукт какого-либо процесса. Хотя производители обычно контролируют то, что выпускают, в отношении данных невозможно отследить все варианты их дальнейшего применения. Данные, созданные для одних целей, могут впоследствии использоваться для других. Важнейшая подгруппа производителей данных – владельцы бизнес-процессов, в рамках которых они создаются. Используются ли данные немедленно или передаются дальше по потоку, владельцы бизнес-процессов оказывают важное влияние на цепочку поставок. Они хорошо знают цели и функции процессов, которыми владеют, и могут вносить в эти процессы изменения, гарантирующие, что производимые данные отвечают требованиям потребителей.

Потребители данных (data consumers)

Потребители данных – это люди и системы, которые используют данные. Слово «потребитель» здесь не совсем уместно, поскольку, как мы знаем, информация не истощается³⁴². Однако оно гораздо лучше отражает суть этой роли в контексте цепочки поставок данных, по сравнению, например, со словом «пользователь».

Брокеры данных (data brokers)

Брокеры данных, также называемые информационными брокерами (information brokers), иногда упускаются из виду в модели производитель – потребитель. В управлении данными они выполняют функции посредников. Брокеры не производят данные, но предоставляют другим возможность их использования. Они похожи на дистрибьюторов промышленных товаров. Информационных брокеров важно распознавать, поскольку они часть цепочки поставок и могут влиять на структуру и содержимое информационных массивов, а также на доступность и своевременность данных (а следовательно, и на их

³⁴⁰ Особенности управления основными данными рассматриваются в главах «Обеспечение доступности и обслуживание данных: основы» (глава 12) и «Управление основными данными: практика внедрения» (глава 13).

³⁴¹ Там же.

³⁴² См. седьмой закон информации, который обсуждался в главе 5.

качество).

Владельцы данных (data owners)

Выработка единого понимания концепции владения данными для большинства организаций – сложная задача. Данные нематериальны и не всегда воспринимаются как актив. Когда они признаются активом, то обычно считаются активом организации, но в итоге такой подход может означать, что о данных никто не заботится. У большинства других организационных активов владельцы, по сути, отсутствуют. Подразделения, управляющие оборудованием, могут нести за него ответственность, но им не владеют. Аналогичным образом, финансовая служба отвечает за денежные активы, которые ей не принадлежат. Активами владеет сама организация. Разница между данными и другими активами заключается в том, что за остальные активы четко определена ответственность (но не владение), и обычно она сосредоточена в одном подразделении или функции внутри организации. Отличие данных, как мы видели, состоит в том, что они перемещаются между подразделениями и функциями.

В то время как блок ИТ безусловно несет ответственность за системы, в которых хранятся и обрабатываются данные, ответственность за эти данные исторически им отрицается. Как правило, ИТ-менеджеры не вникают в содержание информационных ресурсов и не хотят отвечать, если оно не соответствует требованиям бизнеса. Но, поскольку ИТ-подразделения отвечают за системы, они все же в значительной степени контролируют данные в любой из них, а также перемещение данных между системами. Это горизонтальное перемещение вносит еще большую неопределенность в вопросы ответственности, так как по мере движения данных по информационной цепочке они могут быть скопированы или преобразованы в информационные массивы, контролируемые другими системами или командами.

Желание определить владельцев данных можно рассматривать как стремление разрешить указанные проблемы, в частности исправить положение, при котором данные организации не отвечают ее потребностям. Когда сотрудники чувствуют, что данные неполны, противоречивы или неупорядоченны, они ищут владельца, который мог бы лучше контролировать информационные активы.

Как и другими активами организации, данными необходимо управлять. Управление включает в себя знание того, какими данными располагает организация, как их использовать для достижения целей организации и как свести к минимуму любые риски, связанные с их использованием. Данными трудно управлять, поскольку они не только нематериальны, но и очень легко множатся, а также потому, что многие организации не проводят четкой границы между управлением данными и управлением системами, в которых хранятся данные. Часто конфликтные отношения между ИТ-менеджерами и специалистами со стороны бизнеса делают границу между данными и системами еще более размытой. Сбалансировать затраты и выгоды при принятии решений относительно управления данными очень непросто, поскольку сотрудники могут ясно представить себе затраты на ИТ-системы, но им не всегда очевидны выгоды от обрабатываемых в организации данных. Сам собой напрашивается вывод о том, что если ввести роль владельца данных, то будут даны ответы на все возникающие вопросы и управлять данными будет легко.

К сожалению, простого решения сложных задач управления данными не существует. Однако известны эффективные подходы. Один из них – обеспечение внутри организации строгой подотчетности за данные на протяжении всего их жизненного цикла. При этом (так, в частности, считает Лаура Себастьян-Коулман) организациям вовсе не обязательно называть соответствующую сферу ответственности владением данными, но если они считают это полезным, то роли владельцев данных целесообразно ввести³⁴³³⁴⁴.

Распорядители данных (data stewards)

Распорядителем обычно называют лицо, чья работа заключается в управлении собственностью другого лица. Поскольку выше мы обсудили сферу ответственности, которую можно обозначить как

³⁴³ Вопросы владения данными и распределения ответственности тесно связаны с темой руководства данными (data governance), которая подробно рассматривается в главе 10.

³⁴⁴ Там же.

владение данными, логично предположить, что в организации должны быть сотрудники, осуществляющие непосредственное управление информационными активами от имени владельцев данных в интересах организации. Такие сотрудники называются распорядителями данных.

Согласно определению из DMBOK2 распоряжением данными называется деятельность, связанная с несением ответственности и подотчетностью за данные и процессы, обеспечивающие эффективный контроль и использование информационных активов³⁴⁵. Распорядители данных выполняют широкий круг функций и различаются по своей позиции в организации и направлению работы. Основные категории этих специалистов мы рассмотрим в главах 10 («Руководство данными») и 16 («Организационные аспекты управления данными»).

8.5. Эволюция управления данными в организациях и референтные модели

Управление данными рассматривается как отдельная область практики и исследования начиная с тех пор, как компании и правительственные учреждения в начале 1980-х годов стали активно использовать для поддержки своей деятельности базы данных и прикладные системы. За прошедшее время роль данных в организациях существенно изменилась и были накоплены обширные знания, связанные с управлением данными. Одна из особенностей области управления данными – большое количество референтных моделей с представительной базой активных пользователей. Это дает возможность изучить и проанализировать, как референтные модели позволяют накапливать знания в области, имеющей решающее значение для цифровизации³⁴⁶.

Можно выделить три основных этапа развития концепции управления данными (см. табл. 8.1). Они обусловлены технологическим прогрессом и изменениями роли данных. Каждый этап направлен на решение проблем, возникающих в результате этих изменений, и вводит новые подходы к управлению данными, расширяя базу имеющихся знаний.

В таблице 8.2 собраны сведения о наиболее известных референтных моделях управления данными. Для каждой модели указаны организация-поставщик и год ее появления. В списке представлены только те модели, которые имеют практическое значение. Модели концептуального или маркетингового характера в него не вошли.

Из приведенного списка только модели, относящиеся к позициям 1–4, охватывают третий этап развития концепции управления данными. Из них две модели под номером 2 (CDQM и DXM) носят скорее исследовательский характер и довольно сложны для практического применения в большинстве организаций.

³⁴⁵ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

³⁴⁶ Legner C., Pentek T., Otto B. Accumulating Design Knowledge with Reference Models: Insights from 12 Years, Research into Data Management // Journal of the Association for Information Systems, 2020, 21(3): 735–770. DOI: 10.17705/1jais.00618. – URL: https://www.researchgate.net/publication/341684789_Accumulating_Design_Knowledge_with_Reference_Models_Insights_from_12_Years%27_Research_into_Data_Management.

Таблица 8.1.

Эволюция управления данными в организациях (по материалам*)

	Этап 1. Администрирование данных (с 1980-х гг.)	Этап 2. Управление данными, ориентированное на качество (с 1990-х гг.)	Этап 3. Развитие до стратегического управления данными (с 2010-х гг.)
Бизнес-контекст			
Роль данных	Данные как необходимый элемент при разработке приложений и автоматизации бизнес-функций	Данные как необходимый элемент обеспечения общекорпоративных бизнес-процессов и принятия решений	Данные как фактор совершенствования бизнес-моделей и ценностных предложений фирм
Информационные ресурсы	- Базы данных для автоматизированной обработки информации в рамках отдельных организационных функций — например, в системах бухгалтерского учета и системах управления запасами - Структурированные данные	- Интегрированные информационные системы: системы планирования ресурсов предприятия (ERP), компьютеризированное интегрированное производство (CIM) - Хранилища данных, бизнес-аналитика (BI) - В основном внутренние структурированные данные	- Интегрированные и взаимодействующие информационные системы - Озера данных и платформы продвинутой аналитики - Большие объемы структурированных и неструктурированных данных (big data) из внутренних и внешних источников
Проблемы, связанные с данными	Качество модели данных, доступность, повторное использование	Интеграция данных в масштабах организации, качество данных	Ценность и влияние данных на бизнес, нормативно-правовое соответствие, конфиденциальность, безопасность
Ответственность за данные	Администратор базы данных	Владельцы бизнес-процессов, а затем команды по управлению основными данными и бизнес-аналитике	Директор по данным, ученый по данным, аналитик данных

	Этап 1. Администрирование данных (с 1980-х гг.)	Этап 2. Управление данными, ориентированное на качество (с 1990-х гг.)	Этап 3. Развитие до стратегического управления данными (с 2010-х гг.)
Знания по управлению данными			
Взгляд на управление данными	Администрирование данных (в центре внимания — базы данных)	Управление данными, ориентированное на качество (в центре внимания — данные как корпоративный ресурс)	Стратегическое управление данными (в центре внимания — инновации на основе данных)
Подход к управлению	Управление базами данных	Управление ресурсами, управление качеством	Стратегическое управление
База знаний по управлению данными	В основном знания, связанные с базами данных (моделирование данных)	Свод знаний по управлению данными, охватывающий организационные способности (руководство данными [data governance]) и технические способности (интеграцию и архитектуру данных)	Свод знаний по управлению данными, расширяющий организационные и технические способности в отношении получения, хранения и обработки растущего разнообразия и объема данных, а также охватывающий стратегические способности в отношении обеспечения монетизации данных и инноваций на основе данных

* Legner C., Pentek T., Otto B. Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management // Journal of the Association for Information Systems, 2020, 21(3): 735–770.
DOI: 10.17705/1jais.00618. – URL:

https://www.researchgate.net/publication/341684789_Accumulating_Design_Knowledge_with_Reference_Models_Insights_from_12_Years%27_Research_into_Data_Management.

Таблица 8.2.

Референтные модели управления данными (по материалам*)

№	Референтная модель	Тип модели	Авторы
1	DAMA-DMBOK — первая редакция (DAMA, 2009)	Модель способностей	Отраслевой консорциум (включающий экспертов по управлению данными из 53 отделений DAMA на всех континентах)
	DAMA-DMBOK — вторая редакция (DAMA, 2017)	Модель способностей	
2	Модель управления качеством корпоративных данных (Reference Model for Corporate Data Quality Management, CDQM) (Группа исследователей из Европы, 2009)	Модель зрелости и способностей	Научно-промышленный консорциум (более 30 европейских компаний, 3 университета)
	Модель обеспечения высокого качества данных (Data Excellence Model, DXM) (Группа исследователей из Европы, 2017)	Модель зрелости и способностей	
3	Модель зрелости управления данными (Data Management Maturity, DMM) (Институт моделирования зрелости способностей — CMMI, 2014)	Модель зрелости и способностей	Отраслевой консорциум (более 50 авторов, более 70 экспертов-рецензентов, более 80 организаций)
4	Модель оценки способностей по управлению данными (Data Management Capability Assessment Model, DCAM) (Совет по управлению корпоративными данными — EDM Council, 2015)	Модель зрелости и способностей	Отраслевой консорциум (более 200 компаний из финансового сектора)
5	Модель зрелости качества данных (Data Quality Maturity Model) (Совет по повышению эффективности при правительстве США — PIC, 2016)	Модель зрелости	Консорциум правительственных организаций (правительственные учреждения США)
6	Модель зрелости руководства данными (Совет IBM по руководству данными — IBM Data Governance Council, 2007)	Модель зрелости	Отраслевой консорциум (52 компании, ассоциации и правительственные организации, 3 университета)

* Legner C., Pentek T., Otto B. Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management // Journal of the Association for Information Systems, 2020, 21(3): 735–770. DOI: 10.17705/1jais.00618. – URL: https://www.researchgate.net/publication/341684789_Accumulating_Design_Knowledge_with_Reference_Models_Insights_from_12_Years%27_Research_into_Data_Management.

Как уже было отмечено в главе 6, в настоящее время, по мнению ряда специалистов, наиболее полные и ценные с методической точки зрения (а также не зависящие от поставщика соответствующих решений) референтные модели управления данными – это DAMA-DMBOK и CMMI DMM. При этом первая ориентирована на формирование способностей организации по управлению данными, а вторая – на оценку зрелости этих способностей³⁴⁷.

Литература к главе 8

- Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.
- Deng Z. MIS2502: Data Analytics: Semi-structured Data Analytics. Fox School of Business. Temple University, 2019. – URL: <https://slidetodoc.com/mis-2502-data-analyticssemistructured-data-analytics-zhe/>.
- Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.
- Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярчук. – М.: РАНХиГС, 2022.

Глава 9. Управление данными: принципы и структуры

9.1. Методология DAMA-DMBOK

DAMA-DMBOK

В настоящее время наиболее полный и всеобъемлющий методический документ, ориентированный на формирование способностей организации по управлению данными, – DAMA-DMBOK³⁴⁸. Более 30 лет назад была создана Международная ассоциация управления данными (Data Management Association International, DAMA)³⁴⁹, объединяющая профессионалов в области управления данными по всему миру и призванная собирать, систематизировать и пропагандировать лучший опыт. Важная задача DAMA – выработка единой терминологии и рамочных структур (frameworks) знаний и подходов.

В 2009 году вышло первое издание руководства DAMA к своду знаний по управлению данными (DMBOK), заложившее фундамент развития и созревания свода профессий, относящихся к управлению

³⁴⁷ Вопросы оценки зрелости управления данными рассматриваются в следующей главе.

³⁴⁸ Мы уже отмечали в главах 6 и 8, что, по мнению ряда специалистов, сегодня существует фактически только два полноценных комплекта справочных материалов по управлению данными, которые не зависят от поставщика соответствующих решений: DAMA-DMBOK и CMMI DMM (ориентирован на оценку зрелости способностей по управлению данными)*.

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publicatio s, 2017.

³⁴⁹ <https://www.dama.org/>.

данными, а в 2017 году увидело свет второе издание – DAMA-DMBOK2^{350, 351}.

Основное назначение DMBOK – предоставление наиболее полного, точного и актуального введения в дисциплину управления данными:

- выработка общепринятого согласованного представления об областях знаний по управлению данными;
- определение руководящих принципов управления данными;
- предоставление стандартных определений для наиболее часто используемых понятий;
- обзор лучших практик, методов и методик, а также альтернативных подходов;
- краткий обзор общих организационных и культурных вопросов;
- уточнение границ сферы управления данными;
- предоставление рамочных структур управления данными.

Принципы управления данными

Как и в любых других процессах, связанных с менеджментом, в управлении данными должны сбалансированно учитываться как стратегические цели, так и текущие операционные задачи. Для нахождения и соблюдения оптимального баланса рекомендуется следовать определенному своду принципов (рис. 9.1), которые отражают наиболее характерные особенности управления данными и служат проводником в этом процессе.

Для обеспечения последовательной и согласованной реализации этих принципов области управления данными должны быть правильно отструктурированы.

³⁵⁰ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

³⁵¹ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

ПРИНЦИПЫ УПРАВЛЕНИЯ ДАННЫМИ

Эффективное управление данными требует лидерства и приверженности руководства

Данные имеют ценность

- Данные — актив с уникальными свойствами
- Ценность данных может и должна выражаться в экономических терминах

Требования к управлению данными являются бизнес-требованиями

- Управление данными подразумевает управление качеством данных
- Для управления данными необходимы метаданные
- Для управления данными требуется планирование
- Требования к управлению данными должны оказывать определяющее влияние на решения в области информационных технологий

Управление данными требует разнообразных навыков

- Управление данными — кросс-функциональный процесс
- Управление данными требует целостного взгляда на функционирование организации
- Управление данными должно осуществляться с учетом разноплановых перспектив

Управление данными — это управление их жизненным циклом

- Данные различного вида имеют различные характеристики жизненного цикла
- Управление данными включает управление рисками, связанными с данными

Рис. 9.1.

Принципы управления данными*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. — М.: Олимп-Бизнес, 2020.)

Основная рамочная структура DAMA

Основная рамочная структура DAMA состоит из трех элементов: колесо DAMA, шестиугольник факторов среды и контекстные диаграммы, каждый из которых играет свою важную роль.

Колесо DAMA

На рисунке 9.2 представлена круговая диаграмма областей знаний по управлению данными — колесо DAMA (DAMA Wheel). В DMBOK выделено 11 областей знаний по управлению данными³⁵².

1. **Руководство данными (data governance).** Деятельность по осуществлению руководящих и контрольных полномочий (планирование, мониторинг и обеспечение выполнения) в отношении управления информационными активами.

³⁵² В DMBOK области знаний по управлению данными также называются функциями управления данными.

2. Архитектура данных. Определение потребностей организации в данных (безотносительно к структуре), а также разработка и сопровождение основных рабочих описаний (master blueprints) решений по их обеспечению. Использование основных рабочих описаний в качестве руководящих материалов при осуществлении интеграции данных и контроля информационных активов, а также при согласовании инвестиций в области данных с бизнес-стратегией.

3. Моделирование и проектирование данных. Процесс выявления, анализа и формулирования требований к данным с последующим их представлением и распространением в точно определенной форме, называемой моделью данных. Процесс носит итерационный характер и может включать разработку концептуальной, логической и физической моделей.

4. Хранение и операции с данными. Проектирование и реализация решений для хранения, а также сопровождение хранимых данных, с целью получения от них максимальной выгоды на протяжении всего их жизненного цикла.

5. Безопасность данных. Планирование, разработка и осуществление политик и процедур, обеспечивающих должную аутентификацию, авторизацию и доступ пользователей, а также аудит данных и информационных ресурсов организации.

6. Интеграция и интероперабельность данных. Управление перемещением и консолидацией данных как внутри хранилищ, приложений и организаций, так и в рамках обеспечения их взаимодействия.

7. Управление документами и контентом. Планирование, реализация и контроль деятельности по управлению жизненным циклом неструктурированных (или полуструктурированных) данных и информации, представленных в любых форматах и на любых носителях.

8. Справочные и основные данные. Управление совместно используемыми данными, направленное на достижение целей организации, минимизацию рисков, обусловленных избыточностью данных, обеспечение повышения качества данных и снижения затрат на интеграцию данных.

9. Ведение хранилищ данных и бизнес-аналитика. Планирование, внедрение и контроль процессов предоставления данных для принятия решений и информационная поддержка специалистов, участвующих в аналитической деятельности и формировании отчетности.

10. Метаданные. Планирование, организация и контроль деятельности по обеспечению доступа к качественным, интегрированным метаданным.

11. Качество данных. Планирование, организация и контроль выполнения работ по применению стандартных методов управления качеством к данным с целью обеспечения их пригодности к использованию.

Руководству данными отведено центральное место в структуре деятельности по управлению данными, поскольку оно призвано обеспечить согласованность и сбалансированность всех функций. Другие области знаний равномерно распределены вокруг центра колеса DAMA. Все эти компоненты необходимы, чтобы функцию управления данными можно было назвать зрелой, но реализовывать их можно постепенно и в различном порядке, определяемом нуждами организации.

Шестиугольник факторов среды

По каждой области знаний управления данными в DMBOK подробно описаны основные факторы среды (на рисунке 9.2 представлены в виде шестиугольника):

- процессы;
- люди (роли и ответственность);
- технологии.

Шестиугольник факторов среды отражает связи между людьми, процессами и технологиями и служит ключом к прочтению контекстных диаграмм DMBOK (см. далее). В центр помещены цели и принципы, поскольку именно ими необходимо руководствоваться, принимая решения о том, как подойти к исполнению конкретных работ и какие инструменты использовать для эффективного управления данными.

Контекстные диаграммы

В DMBOK применен унифицированный подход к описанию областей знаний по управлению данными, который обеспечивает высокую степень наглядности, – контекстные диаграммы. Контекстная

диаграмма области знаний описывает отдельные ее элементы, включая те, что относятся к людям, процессам и технологиям (отраженным на шестиугольнике факторов среды). За основу построения контекстных диаграмм взят принцип структуризации, применяемый в диаграммах SIPOC (suppliers, inputs, processes, outputs, consumers – поставщики, входы, процессы, выходы, потребители), широко используемых при анализе цепей поставок, о которых мы говорили в главе 7³⁵³. Шаблон контекстной диаграммы приведен на рисунке 9.3.

В контекстных диаграммах центральное место отводится работам, поскольку они дают результаты, удовлетворяющие требованиям тех, кто в этих результатах заинтересован. Каждая контекстная диаграмма начинается с определения и целей области знаний. Работы, обеспечивающие продвижение к целям, помещены в центральной части диаграммы и распределены по четырем фазам – планирование (П), разработка (Р), операции (О) и контроль (К).

1. Работы по планированию (П) определяют стратегический курс и тактику достижения целей управления данными. Работы по планированию повторяются на регулярной основе.

2. Работы по разработке (Р) строятся вокруг жизненного цикла разработки систем (system development lifecycle, SDLC), включающего фазы анализа, проектирования, сборки, тестирования, подготовки и развертывания.

3. Работы по выполнению операций (О) включают мероприятия по поддержке применения, обслуживания и совершенствования систем и процессов, обеспечивающих доступ к данным и их использование.

4. Работы по контролю (К) включают мероприятия по непрерывной поддержке качества данных, а также целостности, надежности и защищенности систем, обеспечивающих доступ к данным и их использование.

Слева («втекают» в работы) указаны входные материалы и их поставщики. Справа («вытекают» из работ) – результаты работ и их потребители. Наконец, участники работ и их роли определены под работами. В нижней части диаграммы перечисляются инструменты, методы и метрики, относящиеся к рассматриваемой области знаний.

Списки в контекстной диаграмме носят иллюстративный и не самый исчерпывающий характер. На практике они могут дополняться другими пунктами и варьироваться по составу в зависимости от специфики организаций. В списки ролей включаются лишь самые важные роли. Каждая организация может адаптировать этот шаблон с целью обеспечения соответствия своим потребностям.

Диаграммы SIPOC помогают отчетливо представить управление информационными активами как процесс формирования цепочки поставок данных. Каждая функциональная область преобразует поступающие на ее вход материалы в продукты, являющиеся вкладом в создание одного из ее звеньев.

³⁵³ Подробнее о диаграммах SIPOC можно посмотреть, например, в «Как наладить процессы» / под ред. К. А. Ткачевой, А. В. Ожаровского. – М.: РАНХиГС, 2021. – URL: <https://lean.cdto.ranepa.ru/>.

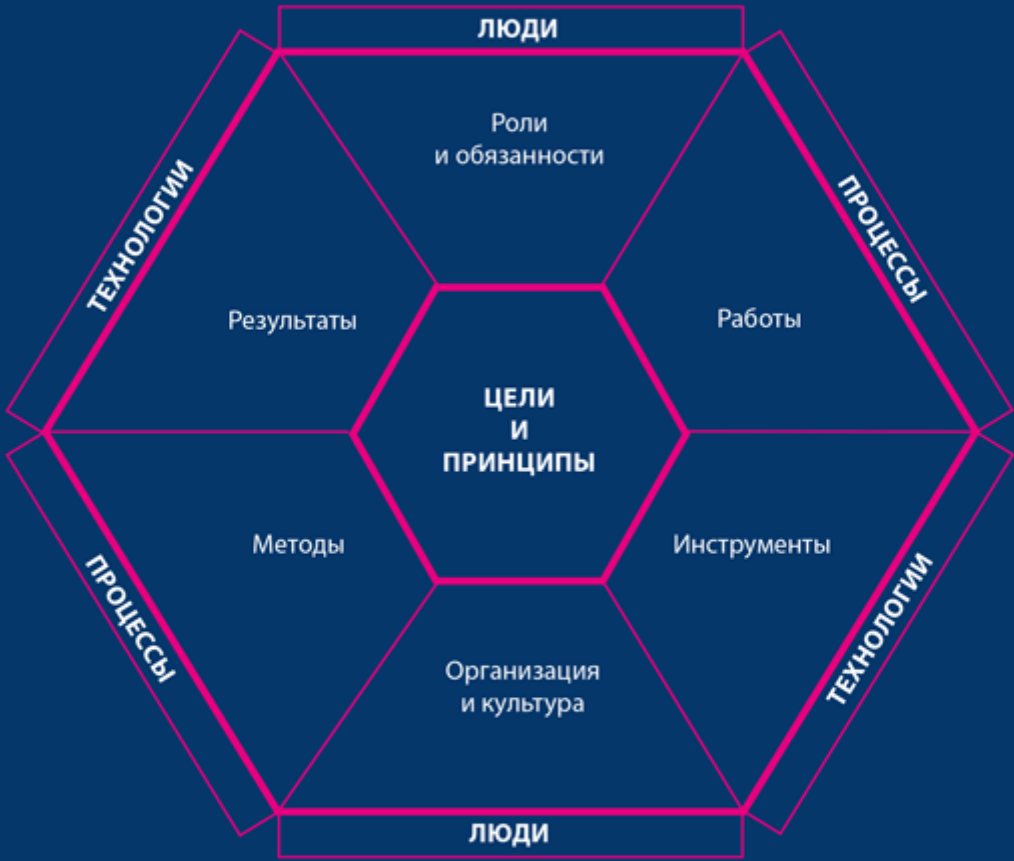


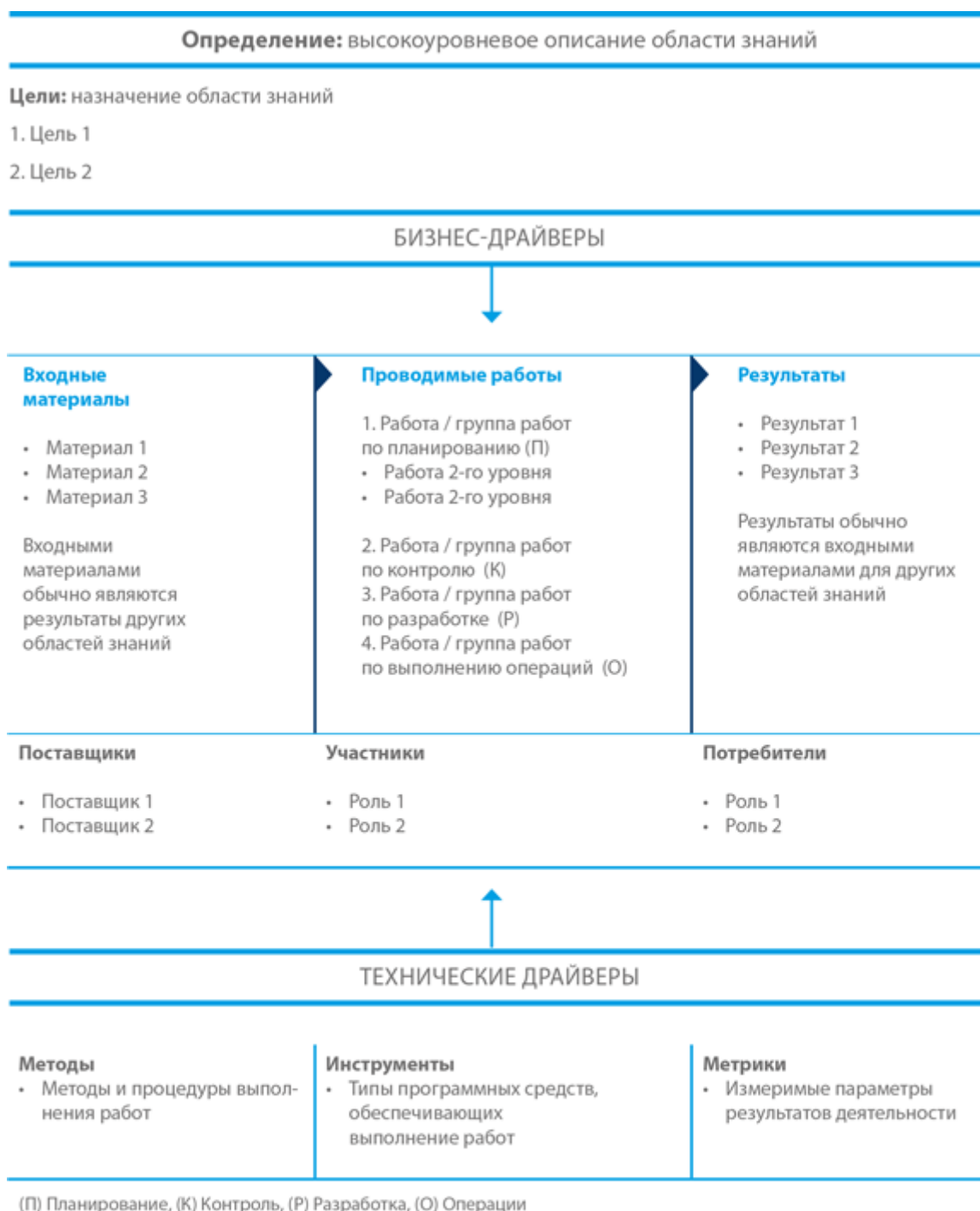
Рис. 9.2.

Колесо DAMA и шестиугольник факторов среды*,**

* Этот рисунок часто используется как логотип DMBOK.

** DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. — М.: Олимп-Бизнес, 2020.)



**Рис. 9.3.**

Контекстная диаграмма области знаний (шаблон)*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Таким образом, элементы основной рамочной структуры DAMA играют следующие роли:

- колесо DAMA отражает состав областей знаний на верхнем уровне;
- на более низком уровне шестиугольник факторов среды выделяет общие структурные компоненты для всех областей;

- контекстные диаграммы позволяют представить детали этих компонентов для каждой области.

В первой части мы уже говорили о том, что метафора цепочки поставок данных позволяет улучшить взаимодействие между отдельными функциями управления данными и более осознанно подойти к проблеме преодоления разрыва между данными и информацией (см. главы 4 и 7). Хотя рассмотренная рамочная структура достаточно полна и подробна, у нее есть недостаток – ни один из трех ее элементов не описывает связи между различными областями знаний. У работающих с ней специалистов нет удобной схемы, наглядно отражающей целостное видение системы управления данными организации и, в частности, позволяющей представить ее в контексте цепочки ценности и цепочки поставок. В результате усилий, направленных на устранение данного пробела, основная структура была дополнена альтернативным представлением областей знаний DMBOK, связанным с жизненным циклом данных.

Представление рамочной структуры DAMA в привязке к жизненному циклу данных

На рисунке 9.4 приведено альтернативное представление колеса DAMA. Этот вариант представления отражает концепцию общей архитектуры областей знаний по управлению данными в виде диаграммы соответствующих функций управления данными и их взаимосвязей.

Диаграмма включает дополнительные детали, разъясняющие содержание некоторых областей там, где это необходимо для более четкого понимания соотношений.

Предлагаемая структура исходит из основной цели управления данными: предоставить организациям возможность извлекать выгоду из их информационных активов так же, как и из любого другого актива. Извлечение выгоды требует управления жизненным циклом данных, поэтому функции управления данными, относящиеся к конкретным фазам жизненного цикла, помещены в центральную часть диаграммы. Начинается все с моделирования и проектирования надежных и качественных данных; затем внедряются процессы и функции, обеспечивающие доступность данных для использования и возможности по их обслуживанию; и, наконец, осуществляется использование данных в различных типах аналитики, за счет чего их ценность повышается.

НАДЗОР: РУКОВОДСТВО ДАННЫМИ

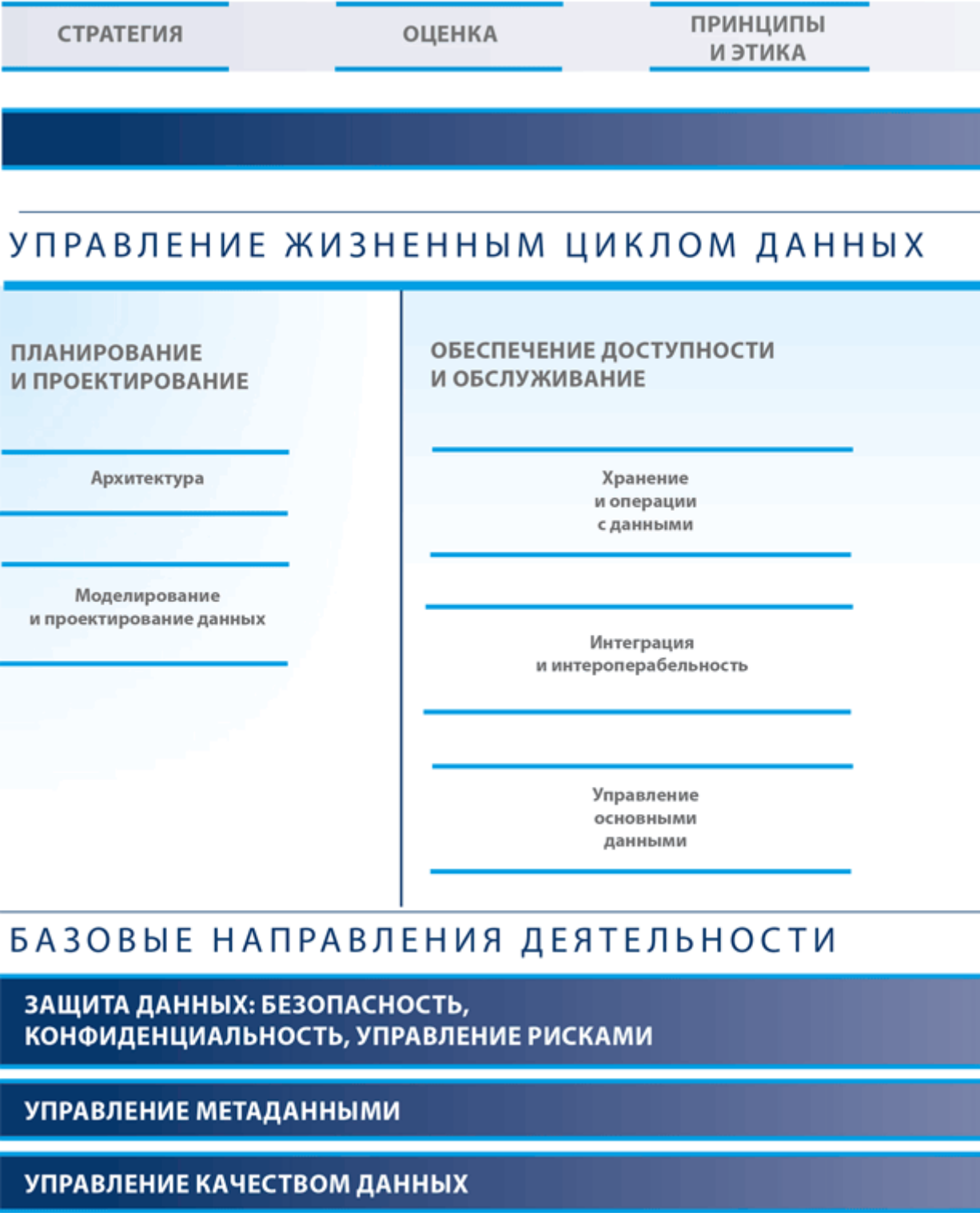


Рис. 9.4. Рамочная структура функций управления данными (в привязке к жизненному циклу)*



* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Отведенный управлению жизненным циклом раздел включает функции по проектированию и

операционные функции (моделирование, архитектура, хранение, обработка и т. п.), необходимые для поддержания традиционных способов применения данных (бизнес-аналитика, управление документами и контентом и т. п.). Кроме того, он учитывает и недавно появившиеся функции, к примеру хранение больших данных, которые необходимы для реализации новых возможностей (наука о данных, предиктивная аналитика и т. п.). Когда организации управляют данными как активом, у них появляется возможность извлечения выгоды из них (монетизация данных) посредством оптимизации бизнеса на основе данных (косвенная монетизация) и (или) их продажи другим организациям (прямая монетизация).

Организации, фокусирующие внимание только на тех функциях, которые напрямую связаны с конкретными фазами жизненного цикла, не смогут извлечь из своих данных такой же большой выгоды, какую они могли бы получить, осуществляя поддержку жизненного цикла с помощью остальных функций управления данными. Эти функции делятся на базовые направления деятельности и деятельность по надзору. Базовые направления деятельности (защита данных, управление метаданными и качеством данных) охватывают весь жизненный цикл данных. Они позволяют реализовывать более качественные проектные решения и облегчают использование данных. Если деятельность по этим направлениям хорошо налажена, стоимость обслуживания данных снижается, потребители данных испытывают к ним больше доверия, а возможности их использования существенно расширяются.

Для успешной поддержки производства и использования данных, а также обеспечения деятельности по базовым направлениям с требуемым уровнем исполнительской дисциплины многие организации устанавливают надзор за управлением данными в форме руководства данными (data governance)³⁵⁴.

Представленная на рисунке 9.4 схема, сильно напоминает структуру цепочки ценностей по Портеру и цепь поставок в соответствии со SCOR-моделью, которые мы обсуждали в главе 7. По сути дела, она отражает модель управления цепочками поставок данных. Как было в сказано в главе 7, хотя на первый взгляд термин «цепочка поставок» ориентирован на поставщика, однако на самом деле цепочки поставок разрабатываются с учетом потребностей клиента. Ориентация на результат позволяет более осознанно подойти к проблеме преодоления разрыва между данными и информацией. Поэтому далее в этой книге мы будем рассматривать функциональные области управления данными именно в соответствии с приведенной на рисунке 9.4 схемой.

9.2. Принципы создания системы управления данными

При создании в организации системы управления данными³⁵⁵ полезно придерживаться следующих принципов³⁵⁶.

1. Принцип владения (принадлежности). Все данные, используемые для обеспечения деятельности организации, должны иметь назначенного владельца, который несет ответственность за надлежащее распоряжение ими.

2. Принцип описания. Все данные должны быть соответствующим образом описаны, чтобы их содержание (и назначение внутри организации) были правильно понимаемы.

3. Принцип обеспечения качества. Все данные должны быть надлежащего качества, соответствующего целям их использования в организации.

4. Принцип обеспечения доступа. Все данные должны быть доступны тем, у кого есть законные основания их использовать. Кроме того, данные должны быть надежно защищены от потери, повреждения или неправильного использования.

5. Принцип совместного использования. Все данные должны быть доступны для обмена с любой организацией или физическим лицом, имеющими соответствующие права, и совместно использоваться

³⁵⁴ Теме руководства данными посвящена следующая глава.

³⁵⁵ Часто в этой связи говорят также о реализации организационной функции управления данными.

³⁵⁶ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

надлежащим образом.

6. Принцип реализации (внедрения и контроля). Должно быть реализовано надлежащее управление всеми категориями данных с учетом специфики каждой категории.

Принципы перечислены в порядке приоритетности. Реализация каждого предыдущего принципа обеспечивает основу для реализации последующих принципов.

Области знаний (функции) управления данными могут быть разбиты на соответствующие функциональные направления, в зависимости от того, к реализации какого принципа они относятся. На рисунке 9.5 эти функциональные направления представлены в виде иерархической пирамиды.



Рис. 9.5.

Иерархия принципов создания системы управления данными*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

Деятельность в рамках того или иного функционального направления имеет приоритет того принципа, который она обеспечивает. Из этого не следует, что нельзя начинать деятельность в рамках какого-либо направления, пока мы не выполним все запланированные мероприятия в рамках более приоритетных направлений. Мероприятия любого направления должны осуществляться параллельно. Просто нужно учитывать, что, если мы не будем уделять должного внимания деятельности по реализации более приоритетных принципов, эффективность мероприятий по реализации остальных принципов будет очень низкой³⁵⁷.

Изложенные принципы, в частности, положены в основу разработки комплекса стандартов по

³⁵⁷ Там же.

управлению данными правительства Абу-Даби (ОАЭ)³⁵⁸. Это один из первых в своем роде всеобъемлющих (с точки зрения охвата областей знаний по управлению данными и этапов жизненного цикла данных) документов в области стандартизации управления данными для правительственных организаций³⁵⁹. Он внедрен и используется в рамках реализации масштабной общеправительственной программы управления данными, охватывающей более 50 государственных ведомств³⁶⁰. Комплекс соответствует рамочной структуре DAMA и обеспечивает поддержку действующей в правительстве Абу-Даби политики управления данными³⁶¹³⁶².

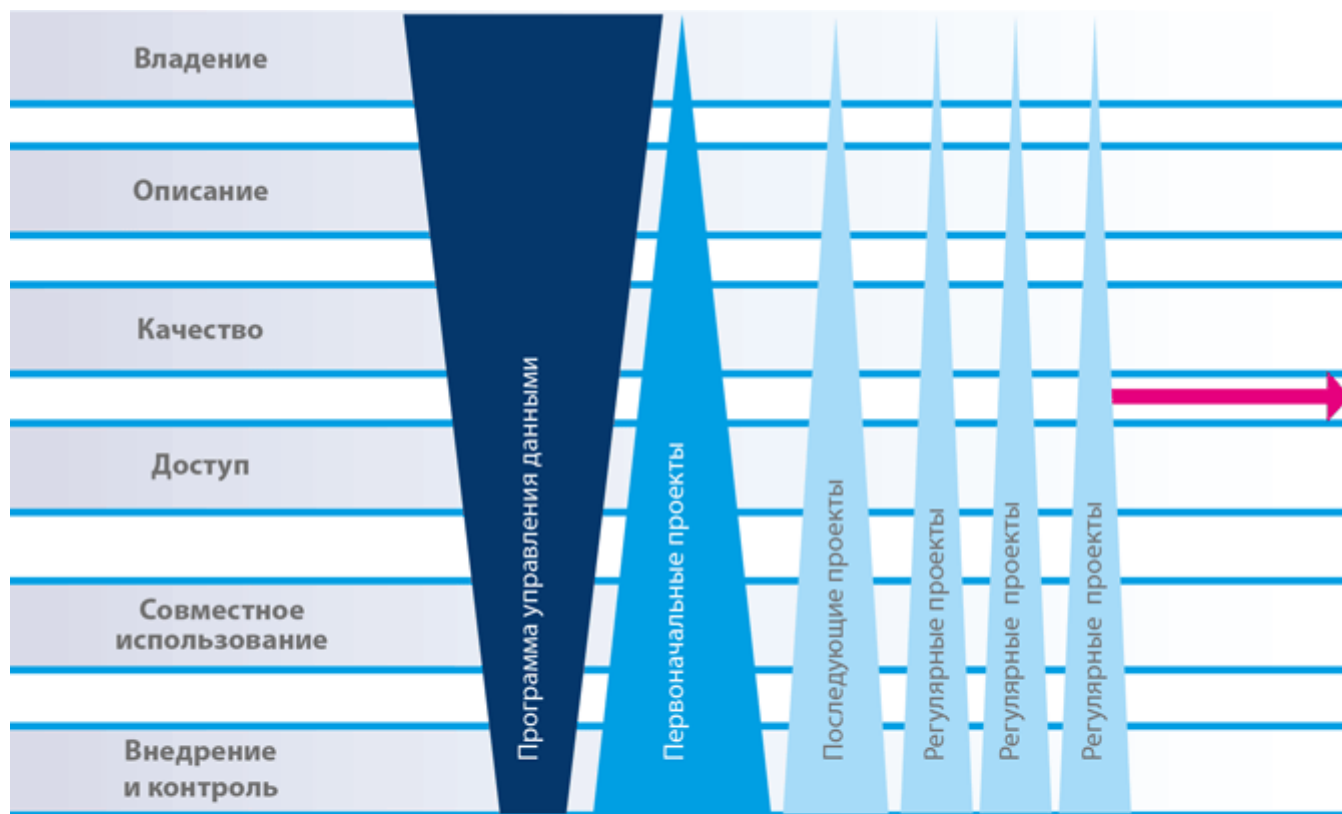


Рис. 9.6.

Распределение усилий по внедрению и выполнению программы управления данными*

³⁵⁸ ADSIC. Abu Dhabi Government Data Management Standards: Version 1.0. Abu Dhabi Systems & Information Centre (ADSIC), 2016. – URL: <https://addata.gov.ae/sites/default/files/AD-Gov-Data-Management-Standards-EN-v1.0.pdf>.

³⁵⁹ Edge J. Abu Dhabi Government Publishes World's First Comprehensive Data Management Standards, 2016. – URL: <https://medium.com/@berrylands/abu-dhabi-governmentpublishes-worlds-first-comprehensive-data-management-standards-3f0e75f0a94f>.

³⁶⁰ ADSIC. Abu Dhabi Government Data Management Program. Abu Dhabi Systems & Information Centre (ADSIC), 2016. – URL: <https://addata.gov.ae/policies>.

³⁶¹ Схожий комплекс стандартов (National Data Management and Personal Data Protection Standards), ориентированный на рамочную структуру DAMA, разработан Национальным офисом управления данными (National Data Management Office, NDMO) Саудовской Аравии для правительственных организаций этой страны (https://www.my.gov.sa/wps/portal/snp/content/dataprotection#header2_1).

³⁶² ADSIC. Abu Dhabi Government Data Management Policy: Version 1.0. Abu Dhabi Systems& Information Centre (ADSIC), 2016. – URL: <https://addata.gov.ae/sites/default/files/AD-Gov-Data-Management-Policy-EN-v1.0.pdf>.

* ADSIC. Abu Dhabi Government Data Management Standards: Version 1.0. Abu Dhabi Systems & Information Centre (ADSIC), 2016. – URL: <https://addata.gov.ae/sites/default/files/AD-Gov-Data-Management-Standards-EN-v1.0.pdf>.

На рисунке 9.6 приведена диаграмма, отражающая распределение усилий по внедрению и дальнейшему выполнению программы управления данными в организации. По вертикали расположены функциональные направления, соответствующие описанным выше принципам. По горизонтали – мероприятия, проводимые в рамках разработки, внедрения и дальнейшего выполнения программы управления данными.

На этапе разработки и внедрения программы управления данными организации необходимо направить основные усилия на начальные мероприятия в верхних функциональных областях (в первую очередь в области, соответствующей принципу владения). Это мероприятия, связанные с реализацией таких необходимых элементов, как организационная система (совокупность организационной структуры и организационного механизма), процессы, инструменты, стандарты.

По мере внедрения постоянно действующей программы управления данными и перехода к выполнению входящих в нее проектов (ИТ– и бизнес-проектов, тесно связанных с работой с данными) на регулярной основе усилия перераспределяются с верхних областей на нижние (прежде всего на область внедрения и контроля управления конкретными категориями данных в конкретных областях бизнеса).

Соответственно, чем лучше будут реализованы элементы верхних функциональных областей на этапе разработки и внедрения программы управления данными, тем меньше дополнительных усилий они потребуют при выполнении регулярных проектов. Это позволит полностью сосредоточиться на достижении основных целей этих проектов – получении максимальной выгоды от работы с данными³⁶³.

9.3. Модель зрелости управления данными

Подход к оценке зрелости способностей (Capability Maturity Assessment, CMM) был разработан в конце 1980-х годов по заказу Министерства обороны США Институтом программной инженерии (Software Engineering Institute, SEI) в Университете Карнеги – Меллона (Carnegie Mellon University, CMU) первоначально для оценки зрелости процессов разработки программного обеспечения. В дальнейшем аналогичные модели появились и для других областей деятельности, в частности для области управления данными^{364, 365}.

В данном случае под зрелостью понимается зрелость процессов, которая, в свою очередь, определяется зрелостью способностей процессов. Под способностью процесса (process capability) понимается характеристика его потенциала для достижения текущих или планируемых бизнес-целей. Модель зрелости процессов была впервые описана Филиппом Кросби (Philip Crosby) – одним из признанных в мире американских авторитетов в области качества – в книге «Качество – бесплатно» (Quality is Free, 1979). Смысл его подхода заключался в следующем: если на предприятии налажены производственные процессы, то уровень качества продукции повышается сам собой, без специальных затрат. Кросби описал пять эволюционных фаз внедрения системы управления качеством. Позже модель зрелости Кросби была адаптирована сотрудниками компании IBM для процессов разработки ПО. В 1986 году они предложили свою концепцию Институту программной инженерии (SEI).

Термин «способности» (capabilities) и важную роль способностей организации в достижении ее

363 Там же.

364 DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

365 Aiken P., Allen M., Parker B., Mattia A. Measuring data management practice maturity: A Community's Self-Assessment. May 2007; Computer 40(4): 42–50. DOI:10.1109/MC.2007.139. – URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.513.2950&rep=rep1&type=pdf>.

стратегических целей мы уже обсуждали, когда говорили о данных как стратегическом ресурсе организации³⁶⁶. Следует заметить, что применительно к вопросам оценки зрелости слово *capabilities* чаще переводится на русский язык как «возможности»³⁶⁷ и для многих отечественных специалистов в сфере ИТ такой перевод более привычен. Исходя из этих соображений, в частности, этот вариант перевода применен и в русском издании DAMA-DMBOK2. Однако, поскольку в этой книге понятие *capabilities* рассматривается не только в контексте оценки зрелости, по отношению к нему применяется более корректный термин – «способности»³⁶⁸.

Модели оценки зрелости управления данными (Data Management Maturity Assessment, DMMA) описываются в терминах продвижения по уровням зрелости, которым поставлены в соответствие характеристики процессов. Когда организация начинает понимать характеристики своих процессов, она может начать повышать их уровень зрелости и внедрять план совершенствования способностей. Она может также измерять степень продвижения и сравнивать себя с конкурентами или партнерами на основе уровней зрелости, выделенных в модели.

С переходом на очередной уровень процесс становится более стабильным, предсказуемым и надежным. Продвижение осуществляется по мере достижения характеристик определенного уровня. Переход на очередной уровень осуществляется, если обеспечиваются соответствующие ему характеристики процесса. При этом ни один уровень не может быть пропущен.

Как правило, рассматриваются следующие уровни (рис. 9.7):

- Уровень 0. Отсутствие возможностей.
- Уровень 1. Начальный (или бессистемный – *ad hoc*): успех зависит от компетенции отдельных сотрудников.
- Уровень 2. Повторяемый: присутствует минимальная дисциплина выполнения процессов.
- Уровень 3. Установленный: введены и используются стандарты.
- Уровень 4. Управляемый: обеспечена возможность измерения характеристик процессов и осуществляется их контроль.
- Уровень 5. Оптимизированный: обеспечена возможность измерения степени достижения целей процессов.

Для каждого уровня описываются критерии оценки характеристик процессов. Например, модель зрелости может включать критерии оценки эффективности процессов, в том числе степень автоматизации. Она может фокусироваться на политиках и контрольных функциях, а также на деталях процессов.

366 См. главу 3.

367 См., например, ГОСТ Р ИСО/МЭК 15504-1-2009 «Информационные технологии. Оценка процессов. Часть 1. Концепция и словарь» *.

* ГОСТ Р ИСО/МЭК 15504-1-2009. Информационные технологии. Оценка процессов. Часть 1. Концепция и словарь.

368 Соответствующий комментарий по поводу вариантов перевода термина *capabilities* приведен в главе 3.

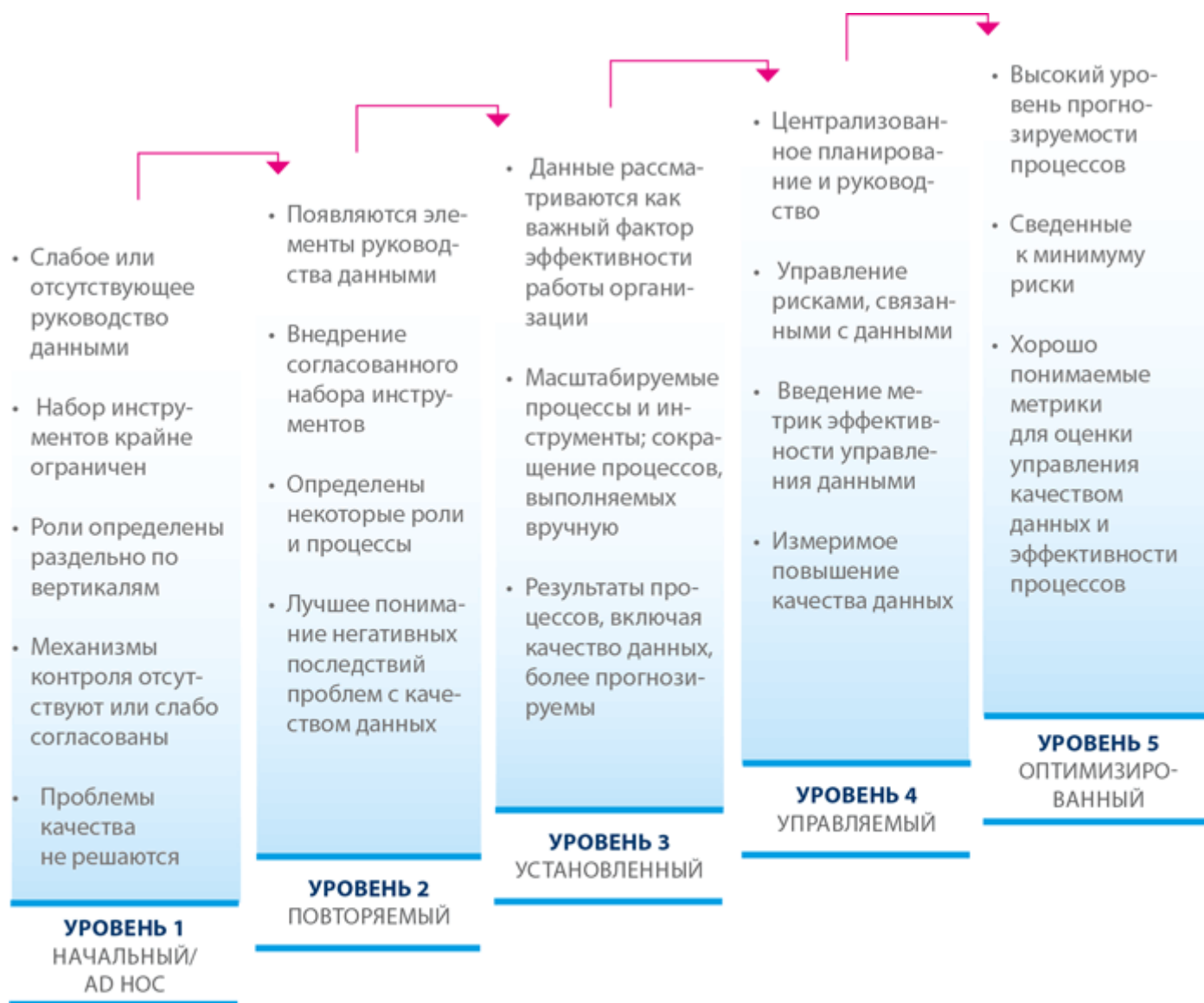


Рис. 9.7.

Пример модели оценки зрелости управления данными*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Оценка зрелости позволяет определить, что работает хорошо, что – недостаточно хорошо и где имеются пробелы. Основываясь на полученных данных, организация может разработать дорожную карту, нацеленную на:

- совершенствование по наиболее важным направлениям, относящимся к процессам, методам, ресурсам и средствам автоматизации;
- обеспечение способностей, которые соответствуют бизнес-стратегии;
- поддержку процессов руководства, которые необходимы для периодической оценки прогресса организации, основанной на характеристиках, заложенных в модель.

Перед началом любого процесса ДММА организации необходимо оценить текущее состояние своих способностей, ресурсов, целей и приоритетов (базовый уровень – baseline). При этом она уже должна обладать некоторой организационной зрелостью – чтобы провести первичную оценку и эффективно отреагировать на ее результаты, определив цели, утвердив дорожную карту и наладив мониторинг прогресса.



Рис. 9.8.

Пример диаграммы оценки зрелости управления данными*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

На рисунке 9.8 отражено возможное визуальное представление результатов экспертной оценки зрелости управления данными по методологии DMMA. Внешний контур задает необходимые для обеспечения конкурентоспособности организации оценки зрелости по всем функциональным областям, а внутренний отражает фактическое положение дел, выявленное по результатам экспертизы. Области с наибольшим разрывом между желаемым и текущим состояниями являются источником наибольших рисков для организации. Подготовка такого отчета очень полезна для определения приоритетов. Для мониторинга достигнутого прогресса могут использоваться повторные экспертизы.

Существует довольно много разнообразных методик оценки зрелости управления данными, разработанных различными организациями. Любая из них предусматривает наличие определенной рамочной структуры из отдельно оцениваемых областей процессов (process areas) управления данными.

Фокус и содержание различных методик сильно варьируются, в зависимости от того, делается упор

на общие вопросы или отраслевую специфику. В то же время во избежание ненужных сложностей, наиболее предпочтительны методики, в которых можно отобразить модель процессов на области знаний по управлению данными.

Из имеющихся сегодня методик наиболее известны, проработаны и подробно описаны две.

1. Модель CMMI-DMM]. Модель зрелости управления данными (DMM) Института моделирования зрелости способностей (Capability Maturity Model Institute, CMMI)³⁶⁹.

2. Модель EDM Council – DCAM. Модель оценки способностей по у, правлению данными (Data Management Capability Assessment Model [DCAM] Совета по управлению корпоративными данными [Enterprise Data Management Council, EDM Council])³⁷⁰.

Модель EDM Council – DCAM ориентирована прежде всего на финансовые организации (в соответствии с основным направлением деятельности EDM Council – отстаивание отраслевых интересов в сфере финансовых услуг), и, хотя ее вполне могут применять организации из других отраслей, все же модель CMMI-DMM гораздо более полная и универсальная.

Со следующей главы мы начнем обсуждение отдельных областей знаний (или функций) по управлению данными. Рассмотрим их роли в формировании цепочек поставок данных и вклад в цепочки ценности. При этом для каждой из функций будут приведены обобщенные характеристики ее уровней зрелости.

Литература к главе 9

• Legner C., Pentek T., Otto B. Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management // Journal of the Association for Information Systems, 2020, 21(3): 735–770. DOI: 10.17705/1jais.00618. – URL:

https://www.researchgate.net/publication/341684789_Accumulating_Design_Knowledge_with_Reference_Models_Insights_from_12_Years%27_Research_into_Data_Management

• Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

Глава 10. Руководство данными

10.1. Руководство данными и его ключевая роль в управлении данными

В главе 6 при обсуждении барьеров на пути развертывания в организации информационных активов мы выделили критически важные факторы успеха – лидерство и приверженность руководства, умноженные на вовлечение всех без исключения сотрудников на всех уровнях организации. В этой главе рассматривается функциональная область, отвечающая непосредственно за работу этого фактора. Она призвана обеспечить реализацию «Сегодняшней повестки дня для совета директоров», о которой мы говорили в главе 7.

Руководство данными (Data Governance, DG) – деятельность по осуществлению руководящих и контрольных полномочий (планирование, мониторинг и обеспечение выполнения) в отношении управления информационными активами организации.

DG – функция управления данными, которая выступает в качестве руководящей всеми остальными его функциями.

³⁶⁹ CMMI Institute. Data Management Maturity (DMM)SM. Build, improve, and measure your Enterprise Data Management Program. ISACA, 2021. – URL: <https://cmmiinstitute.com/dmm>.

³⁷⁰ EDM Council. DCAM Overview, 2021. – URL: https://cdn.ymaws.com/edmcouncil.org/resource/collection/AC65DC50-5687-4942-9B53-3398C887A578/DCAM_Overview_2021_update.pdf.

Цель руководства – обеспечение надлежащего управления данными в соответствии с политиками и лучшими практиками. В то время как главный драйвер управления данными в целом – обеспечение извлечения ценности из информационных активов, функция руководства данными сосредоточена на том, каким образом принимаются решения, касающиеся данных, и как должны функционировать люди и процессы, имеющие к ним отношение.

В предыдущей главе мы сравнили рамочную структуру функций управления данными, отраженную на рисунке 9.4, с моделью управления цепями поставок. Продолжая эту аналогию, трудно представить, что на реальном промышленном предприятии каждый отдел на каждом этапе цепи поставок продукции работает по собственным стандартам качества и по-своему подсчитывает собственные запасы. Подобная ситуация создала бы массу проблем. Чтобы этого не допустить, на предприятии реализуется функция, обеспечивающая аудит и контроль. Наличие такой функции необходимо и в случае цепочки поставок данных. Именно она преобразует управление данными как побочным продуктом или ресурсом в управление ими как активом (рис. 10.1), в то самое управление корпоративной информацией (EIM), о котором мы говорили в главе 7.



Рис. 10.1.

Управление цепями поставок предприятия и цепочками поставок данных*

* Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

Предельно лаконично разъяснить разницу и неразрывную связь между руководством данными и управлением данными можно следующим образом: руководство данными нужно, чтобы «делать правильные вещи» (doing the right things), а управление данными – чтобы «делать вещи правильно» (doing things right)³⁷¹.

Здесь имеется в виду, что руководство обеспечивает определение «правильных вещей» и надзор за их созданием или выполнением, а управление занимается непосредственно их созданием или выполнением (рис. 10.2).

В реальной практике руководство и управление данными тесно переплетены и добиться их идеального разделения не всегда возможно (из-за дефицита кадров и т. п.). Поэтому, чтобы лучше усвоить и постоянно учитывать разницу между руководством и управлением данными, Джон Лэдли рекомендует использовать введенную им хорошо запоминаемую модель V руководства (governance V)(рис. 10.3).



Рис. 10.2.

Руководство данными и управление данными*

³⁷¹ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

* Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.



Рис. 10.3.

У руководства*

* Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

10.2. Стратегия работы с данными

В первой части книги мы уже говорили о том, что в современной организации данные в качестве важнейшего актива играют ключевую роль в обеспечении выполнения ее стратегии. Поэтому важно следовать определенным установкам, которые обеспечивали бы максимальное согласование работы с данными со стратегическими целями бизнеса. Для этого организация создает стратегию работы с данными (data strategy). Это мероприятие стоит первым пунктом в «Сегодняшней повестке дня для совета директоров», которую мы обсуждали в главе 7.

Стратегия работы с данными предусматривает высокоуровневые планы использования информации для получения конкурентных преимуществ и реализации целей организации. Она должна исходить из понимания информационных потребностей, заложенных в бизнес-стратегию, а именно: какие данные нужны организации, как она будет получать эти данные, управлять ими, обеспечивать их надежность и достоверность на протяжении всего жизненного цикла, каким образом их использовать.

Стратегия также должна предусматривать меры, направленные на решение известных проблем в области управления данными.

Как правило, стратегия работы с данными представляет собой документ достаточно высокого уровня, фиксирующий основные цели и принципы работы с данными, а также основные пути ее реализации³⁷², поэтому обычно она дополняется поддерживающей ее более детальной стратегией управления данными (data management strategy)³⁷³.

Согласно DMBOK стратегия управления данными должна включать следующие компоненты:

³⁷² В качестве примера можно привести Стратегию работы с данными Министерства обороны США *.

* Do D. DoD Data Strategy. U.S. Department of Defense (DoD), 2020. – URL: [https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF](https://media.defense.gov/2020/Oct/08/2002514180/1/-1/0/DOD-DATA-STRATEGY.PDF).

³⁷³ Следует обратить внимание на достаточную распространенность случаев нестрогого употребления терминов data strategy и data management strategy. Часто разные авторы вкладывают в них разное содержание. Нередко в одном источнике они могут использоваться одновременно как синонимы.

- убедительно изложенное видение управления данными;
- краткое экономическое обоснование стратегии с избранными примерами (business cases);
- руководящие принципы, ценности и перспективы управления;
- миссию и долгосрочные цели по основным направлениям управления данными;
- предлагаемые показатели успешности управления данными;
- краткосрочные (на 12–24 месяца) задачи программы управления данными, которые должны быть конкретными, измеримыми, практически значимыми, реалистичными и привязанными к точным срокам (согласно принципу SMART: specific, measurable, actionable, realistic, time-bound);
- описания ролей и организационных систем управления данными, включая распределение обязанностей и прав принятия решений;
- описания компонентов и инициатив программы управления данными;
- приоритетную программу работ с объемами и сроками выполнения;
- первоначальный вариант дорожной карты реализации с разбивкой по проектам и мероприятиям.

Иногда выделяются отдельные стратегии по областям управления данными: стратегия руководства данными, стратегия в области качества данных, стратегия в области метаданных, стратегия в области бизнес-аналитики и т. п. Однако многие авторитетные специалисты считают, что предпочтительнее рассматривать единую стратегию управления данными как сумму всех перечисленных стратегий. Это позволит избежать чрезмерных затрат на планирование из-за необходимости координации множества отдельных программ. Особенно в условиях недостаточной зрелости соответствующих областей управления данными³⁷⁴. В DMBOK так же указано, что стратегия управления данными должна охватывать все области знаний, входящие в рамочную структуру DAMA и имеющие отношение к организации.

Руководство данными играет важную роль в обеспечении соответствия между стратегией работы с данными и стратегией организации. Взаимоотношения между стратегией организации, стратегией работы с данными и руководством данными отражены на рисунке 10.4.

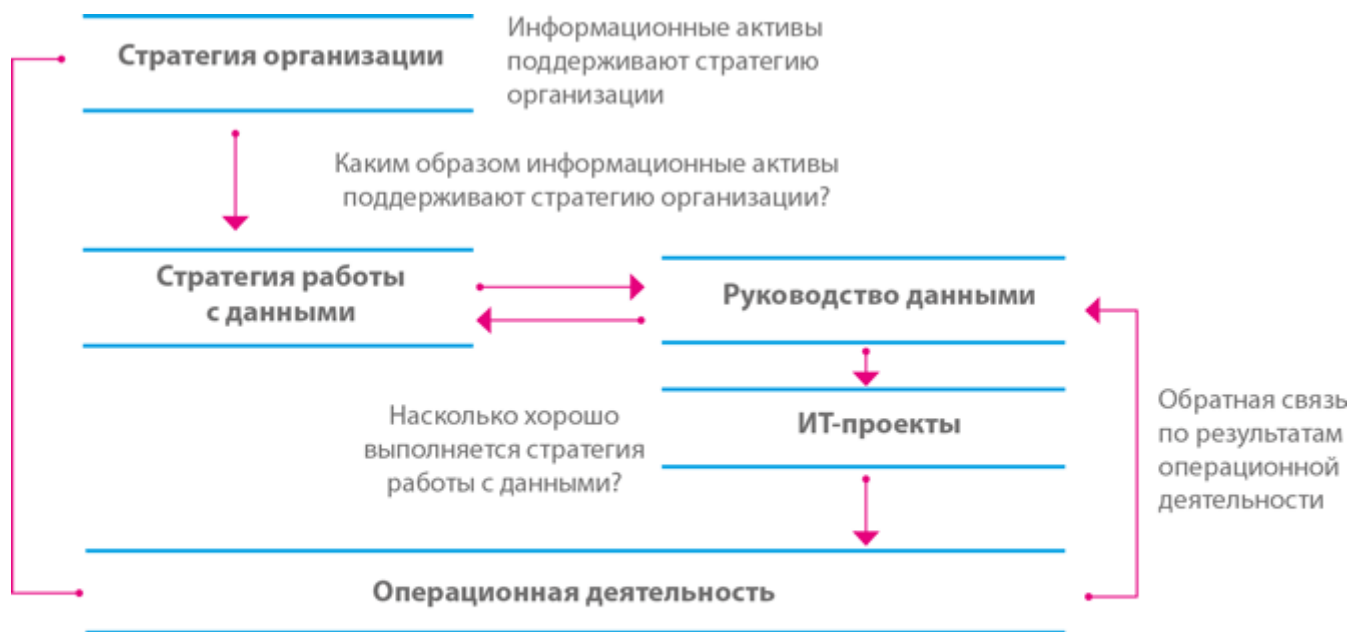


Рис. 10.4.

Стратегия организации, стратегия работы с данными и руководство данными*

* Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

³⁷⁴ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publications, 2017.

Стратегия организации задает основные направления и контекст для стратегии работы с данными. Руководство данными в свою очередь поддерживает реализацию этой стратегии, оказывая влияние на связанные с управлением данными аспекты ИТ-проектов. ИТ-проекты в свою очередь обеспечивают предоставление необходимой информации сотрудникам и партнерам организации.

В главе 6 мы говорили о семи «смертных грехах». В завершение этого раздела нужно заметить, что Питер Айкен рассматривает их прежде всего в контексте реализации стратегии работы с данными. Поэтому в рамках проведения соответствующих мероприятий (от инициирования работы по созданию стратегии до ее внедрения и дальнейшего обновления) следует уделять поднятым Айкеном вопросам самое пристальное внимание.

10.3. Программа руководства данными

Руководство данными – не разовое мероприятие. Для его осуществления требуется постоянно действующая в организации программа, нацеленная на извлечение максимальной выгоды из имеющихся в ее распоряжении данных при одновременной минимизации рисков, связанных с этими данными³⁷⁵.

Программа руководства данными позволяет организации быть «управляемой на основе данных» благодаря наличию стратегии, подкрепленной принципами, политиками и практиками распоряжения данными, которые обеспечивают своевременное выявление и использование возможностей для извлечения выгоды из имеющихся данных. Программа должна также предусматривать привлечение процессов управления организационными изменениями для обучения организации и поощрение поведения, ориентированного на стратегическое использование данных.

Общее содержание и ключевые направления программы руководства данными будут зависеть от потребностей конкретной организации, но большинство программ включает следующие компоненты^{376, 377}:

1. Стратегия. Определение, доведение до исполнителей и управление реализацией стратегии работы с данными и стратегии руководства данными (если она разрабатывается отдельно от стратегии работы с данными).

2. Политика. Определение и обеспечение соблюдения политик в отношении управления, доступа, использования, безопасности и качества данных и метаданных. Разработка политики в области работы с информацией – второй пункт в «Сегодняшней повестке дня для совета директоров», которую мы обсуждали в главе 7. Повестка выделяет основные аспекты управления информацией, которые критически важны для деятельности организации и должны быть отрегулированы в первую очередь.

3. Стандарты и качество. Определение и обеспечение соблюдения стандартов в области качества данных и архитектуры данных.

4. Надзор. Практическое осуществление наблюдения, аудита и исправление выявленных недостатков в ключевых аспектах обеспечения качества, соблюдение политики и управления данными.

5. Нормативно-правовое соответствие (compliance). Обеспечение соблюдения организацией нормативно-правовых требований в отношении данных.

6. Управление проблемными вопросами. Выявление, определение, эскалация и разрешение проблемных вопросов, связанных с безопасностью данных, доступом к данным, качеством данных, нормативно-правовым соответствием, владением данными, политикой, стандартами, терминологией и процедурами руководства данными.

7. Проекты по управлению данными. Поддержка усилий, направленных на развитие практик

³⁷⁵ О необходимости программного подхода к управлению данными мы уже говорили в главе 6, когда обсуждали семь «смертных грехов».

³⁷⁶ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

³⁷⁷ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

управления данными.

8. Оценка информационных активов. Введение стандартов и процессов, позволяющих согласованно определять ценность информационных активов для бизнеса³⁷⁸.

10.4. Распоряжение данными

Распоряжение данными (data stewardship) – деятельность, связанная с несением ответственности за данные и процессы, обеспечивающие эффективный контроль и использование информационных ресурсов. Распорядители данных (data stewards) осуществляют практическую реализацию функции руководства данными.

В большинстве случаев практика распоряжения данными сосредоточена на следующих аспектах^{379, 380}:

1. Создание ключевых метаданных и управление ими. Определение бизнес-терминологии, допустимых значений данных и других важных метаданных и управление ими. Распорядители данных часто отвечают за бизнес-гlossарий организации, который становится системой записи для бизнес-терминов, относящихся к данным.

2. Документирование правил и стандартов. Определение и документирование бизнес-правил, стандартов данных и правил качества данных. Представления в отношении данных, используемые для определения данных высокого качества, часто формулируются в виде правил, которые скрыты в бизнес-процессах, создающих или потребляющих данные. Распорядители данных помогают выявить эти правила, подтвердить их единое понимание в рамках организации и убедиться, что они применяются соответствующим образом.

3. Управление проблемными вопросами в области качества данных. Распорядители данных часто принимают активное участие в выявлении и разрешении проблемных вопросов, обусловленных недостаточным качеством данных, или содействуют их разрешению.

4. Оперативная деятельность по руководству данными. Распорядители данных несут ответственность за обеспечение того, чтобы политика и инициативы в области руководства данными соблюдались ежедневно по каждому проекту. Они должны влиять на принимаемые решения, чтобы обеспечить такое управление данными, которое способствовало бы достижению общих целей организации.

5. Осуществление текущей деятельности по руководству данными. Распорядители данных отвечают за обеспечение постоянного соблюдения политик и поддержку инициатив в области руководства данными в рамках всех реализуемых проектов. Они также должны способствовать принятию решений по управлению данными, создающих условия для достижения организацией своих целей.

Конкретные роли, которые могут выполнять в организации распорядители данных, мы рассмотрим в главе 16 «Организационные аспекты управления данными».

Особенности перевода терминов Data Governance и Data Stewardship на русский язык

При переводе отраслевой терминологии приходится искать компромисс между стремлением использовать привычные и понятные русскоязычные термины и сохранением терминологической стройности, связности и лаконичности оригинального текста. Это в полной мере относится к переводу текста DAMA-DMBOK2.

По отношению к термину Data Governance наиболее приемлемое решение – введение семантического неологизма (сочетания терминов, которое звучит не совсем по-русски, но по смыслу наиболее точно соответствует оригиналу) «руководство данными».

³⁷⁸ Модели оценки информационных активов мы рассматривали в главе 5.

³⁷⁹ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

³⁸⁰ Plotkin D. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance: 2nd Edition. Academic Press, 2020.

Одним из весомых аргументов в пользу этого варианта перевода является пояснение, приведенное в тексте DMBOK2: «Слово “руководство” (governance) образовано от глагола “руководить” (govern), который в данном случае ключевой. Смысл руководства данными проще всего понять на примере политического руководства. В отношении данных предусматриваются функции, подобные законодательным (определение политик, стандартов и корпоративной архитектуры данных), судебным (управление проблемными вопросами и эскалация) и исполнительным (защита и обслуживание, выполнение обязанностей по администрированию). Для лучшего управления рисками большинство организаций выбирают представительную форму руководства данными, обеспечивающую учет мнений всех заинтересованных сторон».

Таким образом, слово «руководство» наиболее **передает смысл, заложенный в оригинале**. При этом следует заметить, что более корректно переводить термин Data Governance как «руководство в области управления данными», но с учетом высокой частоты его использования и наличия многих производных от него терминов такой вариант был бы крайне неудобен.

Подробные пояснения приведены в DMBOK2 и относительно термина Data Stewardship: «Деятельность, связанная с несением ответственности и подотчетностью за данные и процессы, обеспечивающие эффективный контроль и использование информационных активов, чаще всего обозначается термином “распоряжение данными” (data stewardship). Распорядитель (steward) – это лицо, чья работа заключается в управлении собственностью другого лица. Распорядители данных управляют информационными активами от имени других лиц и в интересах организации. Часто по отношению к тем, кто выполняет функции, подобные распорядительским, применяют такие термины-синонимы, как “хранитель” (custodian) или “попечитель” (trustee)»³⁸¹.

10.5. Организационная система руководства и управления данными

Вопросы построения организационной системы работы с данными вынесены в третий пункт «Сегодняшней повестки дня для совета директоров» (см. главу 7). Чтобы обозначить основные акценты, приведем его здесь еще раз.

«Большую часть повседневных обязанностей по управлению информацией нужно делегировать сотрудникам по всей организации, при этом их выполнение следует подкрепить четкой операционной структурой (operational framework) зон ответственности руководителей, команд и отдельных работников, прозрачной на всех уровнях и утвержденной советом директоров. Она будет залогом эффективного руководства данными и выполнения бизнес-планов (дорожных карт), определяющих инициативы в области управления информацией».

На рисунке 10.5 отражена обобщенная организационная система руководства и управления данными³⁸².

В части руководства данными можно выделить следующие типичные комитеты и другие органы.

1. Управляющий комитет по руководству данными (Data Governance Steering Committee). Высший орган руководства данными в организации, отвечающий за надзор, поддержку и финансирование DG. Представляет собой кросс-функциональную группу руководителей высшего звена. Обычно утверждает объемы финансирования деятельности по руководству данными и работ, поддерживаемых со стороны DG (в соответствии с рекомендациями совета по руководству данными [Data Governance Council, DGC] и директора по данным [Chief Data Officer, CDO]). Этот комитет в свою очередь может быть подконтролен также относящемуся к высшему звену комитету по финансированию или другим комитетам, созданным в рамках отдельных инициатив.

2. Совет по руководству данными (Data Governance Council, DGC). Управляет инициативами в области DG, например разработкой политик или метрик, разрешением проблемных вопросов и эскалацией. Состав формируется из руководителей и ответственных сотрудников в соответствии с используемой

381 Скворцов Н. DAMA-DMBOK2: трудности перевода // Открытые системы. СУБД, 02, 2020. – URL: <https://www.osp.ru/os/2020/02/13055423>.

382 Здесь под организационной системой понимается совокупность организационной структуры и организационного механизма. В англоязычных источниках для обозначения этого понятия часто используется слово organization.

операционной моделью.

3. Офис по руководству данными (Data Governance Office, DGO). Ведет текущую работу в части определений данных и стандартов по управлению данными корпоративного уровня во всех функциональных областях управления данными. Формируется из сотрудников (распорядителей данных) с координирующими ролями.

4. Команды по распоряжению данными. Заинтересованные группы сотрудников, фокусирующиеся на одной или нескольких предметных областях или проектах и сотрудничающие или осуществляющие взаимные консультации с проектными командами по вопросам определения данных или стандартов по управлению данными, относящимся к области интересов. Состоят из распорядителей данных со стороны бизнеса или ИТ, а также из аналитиков данных.

5. Локальные комитеты (советы) по руководству данными. В крупных организациях могут формироваться комитеты или советы по руководству данными на уровне отдельных дивизионов или департаментов, работающие при содействии и под наблюдением корпоративного DGC. Небольшим организациям лучше избегать подобных сложностей.



Рис. 10.5.

Руководство и управление данными: организационная система*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

В главе 6 мы рассматривали семь «смертных грехов» в области работы с данными. Сразу за первым

«грехом» – непониманием основ дата-центричного мышления – следует второй – отсутствие квалифицированного руководства и лидерства в области данных. Большая часть ответственности за искупление этого «греха» лежит на директоре по данным. Мы проводили аналогию между его ролью и ролью финансового директора. В большинстве организаций в обязанности CDO входит:

- определение организационной стратегии в области данных;
 - согласование требований к данным с имеющимися ИТ– и бизнес-активами;
 - определение и утверждение стандартов, правил и процедур руководства данными;
 - консультационная поддержка (и при необходимости обслуживание) инициатив бизнеса, зависящих от данных;
 - донесение до понимания всех значимых сторон внутри и вне организации важности соблюдения принципов качественного управления данными;
 - обеспечение надзора за использованием данных в целях бизнес-аналитики^{383, 384, 385}.
- Обсуждение организационных аспектов работы с данными мы продолжим в главе 16.

10.6. Контекстная диаграмма области знаний и уровни зрелости функции «Руководство данными»

Контекстная диаграмма области знаний «Руководство данными» приведена на рисунке 10.6.

³⁸³ Aiken P., Gorman M. The Case for the Chief Data Officer: Recasting the C-Suite to Leverage Your Most Valuable Asset. Morgan Kaufmann Publishers Inc., 2013.

³⁸⁴ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

³⁸⁵ Soares S. The Chief Data Officer Handbook for Data Governance. MC Press, 2014.

РУКОВОДСТВО ДАННЫМИ

Определение: деятельность по осуществлению руководящих и контрольных полномочий, а также обеспечению совместного принятия решений (планирование, мониторинг и обеспечение выполнения) в отношении управления информационными активами

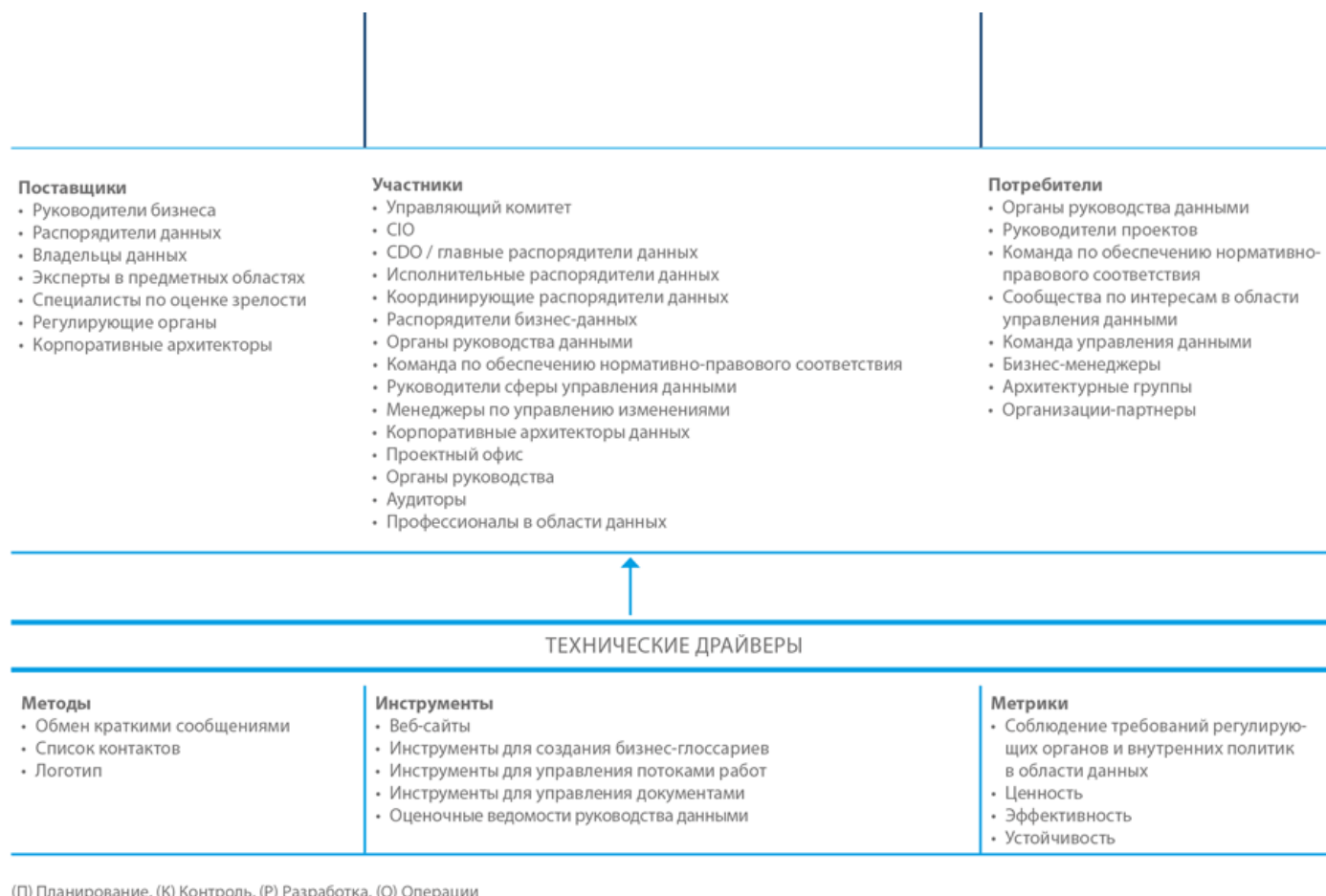
Цели:

1. Создание в организации возможностей для управления данными как активом
2. Определение, утверждение, доведение до всеобщего сведения и внедрение принципов, политик, процедур, метрик, инструментов и обязанностей в сфере управления данными
3. Мониторинг и руководство обеспечением соблюдения политики, использованием данных и деятельностью по управлению

БИЗНЕС-ДРАЙВЕРЫ



Входные материалы	Проводимые работы	Результаты
• Стратегии и цели бизнеса • Стратегии и цели в области ИТ • Стратегии в отношении данных и управления данными • Политика и стандарты организации • Результаты оценки бизнес-культуры • Результаты оценки зрелости данных • ИТ-практики • Нормативно-правовые требования	1. Определение задач и функций руководства данными в организации (П): • Разработка стратегии руководства данными • Проведение оценки готовности • Выявление возможностей/угроз и согласование с бизнесом • Создание точек взаимодействия внутри организации 2. Разработка стратегии руководства данными (П): • Определение операционной рамочной структуры руководства данными • Выработка целей, принципов и политик • Поддержка проектов в области управления данными • Внедрение практики управления организационными изменениями • Внедрение практики управления проблемными вопросами • Оценка требований по нормативно-правовому соответствию 3. Внедрение руководства данными (О): • Поддержка стандартов и процедур • Разработка бизнес-гlossария • Координация взаимодействия с архитектурными группами • Оказание содействия в финансовой оценке данных 4. Внедрение распоряжения данными (К, О)	• Стратегия руководства данными • Стратегия работы с данными • Стратегическая карта развития бизнеса / руководства данными • Принципы, политики и процессы руководства данными • Операционная рамочная структура • Дорожная карта и стратегия внедрения • Операционный план • Бизнес-гlossарий • Оценочная ведомость руководства данными • Веб-сайт DG • План коммуникаций • Признанная оценка стоимости данных • Зрелые практики управления данными



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 10.6.

Контекстная диаграмма. Руководство данными*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

В соответствии с рамочной структурой функций управления данными, представленной на рисунке 9.4, функция руководства данными задействована на всех этапах их жизненного цикла. Схема на рисунке 10.7 отражает распределение деятельности в области DG по отдельным этапам.

В предыдущей главе мы упоминали о комплексе стандартов по управлению данными правительства Абу-Даби, который широко используется в государственных ведомствах ОАЭ. Это один из первых в своем роде всеобъемлющих документов в области стандартизации управления данными для правительственных организаций. Поэтому, думаем, что в завершение обсуждения руководства данными небезынтересно было бы узнать, какие элементы DG выделяются в качестве ключевых разработчиками комплекса (см. рис. 10.8).



Рис. 10.7.

Распределение деятельности в области руководства данными по этапам их жизненного цикла*

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

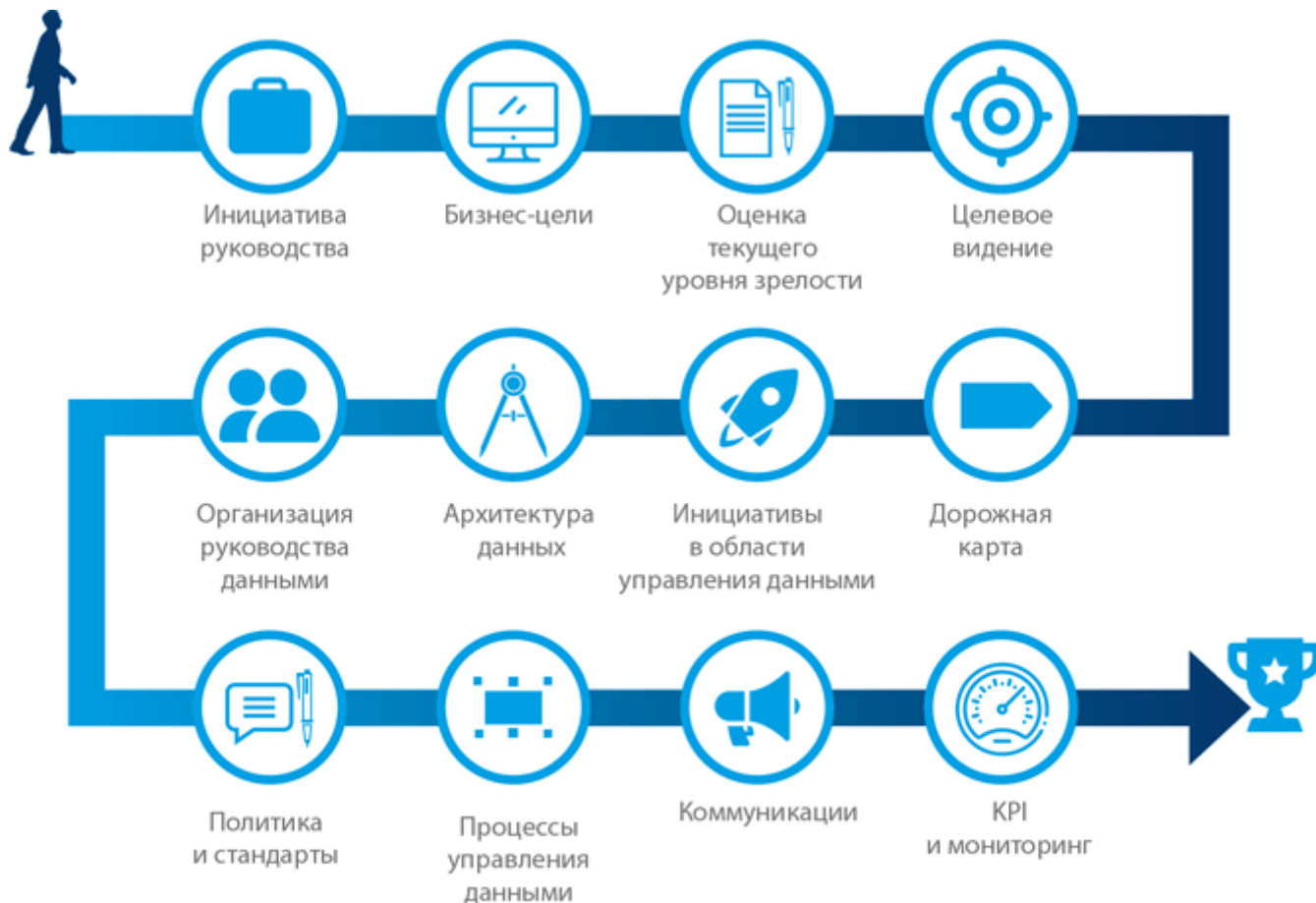


Рис. 10.8.

Ключевые шаги при внедрении руководства данными*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

Каждый из представленных на рисунке 10.8 элементов необходим для внедрения в организации эффективной функции руководства данными. Хотя они выстроены в виде рекомендуемой последовательности шагов, из этого не следует, что проводить работы в области DG нужно исключительно в указанном порядке. Мероприятия могут осуществляться в соответствии с частными приоритетами конкретной организации.

На рисунке 10.9 представлены обобщенные характеристики уровней зрелости функции «Руководство данными».

Начиная со следующей главы мы переходим к обсуждению областей знаний (функций), которые, в отличие от руководства данными, распространяются не на весь жизненный цикл данных, а на отдельные его фазы – планирование и проектирование, обеспечение доступности и обслуживание, использование и расширение возможностей применения (см. рис. 9.4). В завершение обсуждения каждой из этих функций будет проводиться анализ ее влияния на ценность данных. Для руководства данными мы такого анализа не проводим. В главе 7, сравнивая элементы референтной модели управления цепями поставок (SCOR-модель) с цепочкой поставок данных, мы отметили, что руководство данными (вместе с управлением безопасностью данных, метаданными и качеством данных) соотносится с группой процессов «Предоставлять возможность». Поэтому его влияние на ценность вполне понятно. Оно состоит в первую очередь в максимальном содействии увеличению эффективности остальных функций, повышающих ценность данных на отдельных фазах их жизненного цикла.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Управляющий комитет (группа) по руководству данными активно проводит работу по оценке деятельности (своей и подразделений) в области DG на основе набора ключевых показателей эффективности (KPI), при этом подразделения отчитываются о соблюдении практик и стандартов. В качестве средства обеспечения улучшений в отношении персонала, процессов и технологий используется архитектурный подход, ориентированный на повторное использование решений и результатов (строительных блоков) по всей организации.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Введены в действие корпоративные политики и стандарты, устанавливающие стратегическую линию в отношении руководства данными для всех подразделений. Существует общеорганизационный управляющий комитет (группа) по руководству данными, для которого определены роли и обязанности. В подразделениях присутствует понимание процессов в рамках деятельности по руководству данными. Корпоративная стратегия в области архитектуры задает общее направление для всех связанных с данными проектов. Для обеспечения лучшей согласованности проектов со стратегией используются контрольные точки. Процессы и роли, связанные с данными, четко определены и понятны сотрудникам по всей организации.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Имеются свидетельства применения документированных практик руководства данными, но в основном на уровне проектов или подразделений. Архитектурные практики также документируются на уровне подразделений. Нет гарантии, что эти практики являются достаточно всеобъемлющими или что существует возможность их согласованного внедрения в рамках нескольких подразделений или организации в целом. Нет никаких свидетельств проявления на корпоративном уровне мышления, ориентированного на руководство данными.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации отсутствуют свидетельства деятельности в области DG. Вопросы стратегического управления данными не рассматриваются. Практики в области архитектуры (систем и данных) не регламентированы. Мало внимания уделяется возможностям повторного использования и стандартизации на корпоративном уровне. Хотя некоторые определения могут быть согласованы, они, как правило, используются применительно к отдельным проектам.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 10.9.

Обобщенные характеристики уровней зрелости функции «Руководство данными»*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.



ПРАКТИЧЕСКИЙ ПРИМЕР

Начнем рассмотрение обсуждаемых подходов к организации работы с информационными активами на примере нашей вымышленной телекоммуникационной компании «Телеком Дубль», которую мы представили читателям в предисловии.

Чтобы немного охарактеризовать стратегические задачи «Телеком Дубль», рассмотрим

такую важную группу ее клиентов, как частные лица, которые в период пандемии COVID-19 стали активнее пользоваться сервисами компании. Особый спрос получил продукт по проведению видеоконференций. Этот рынок резко вырос, и, чтобы привлечь как можно больше клиентов, компания незамедлительно начала предлагать новые возможности, повышающие удобство пользователей. Отдельной стратегией было понижение порога вхождения новых клиентов. Новые сервисы предусматривали обратную связь для пользователей, приложение для сбора аналитики по эксплуатации сервиса, упрощение регистрации новых клиентов, дополнительные средства по защите конфиденциальных данных, возможность создания групповых звонков, расширение спектра ролей пользователей, организацию комнат внутри групповых видеоконференций.

«Телеком Дубль» работает не только с физическими, но и с юридическими лицами – корпоративными клиентами. Подразделение продаж компании постоянно запрашивает аналитику по таким пользователям: чтобы свести к минимуму их отток и повысить объемы продаж. Один из основных используемых при этом инструментов – аналитическая система, построенная как на внутренних, так и на внешних данных. С одной стороны, ведется сбор внутренней информации по объемам расходов, количеству сотрудников, статистике пользования продуктами. С другой – эту информацию нужно обогатить внешними данными: сведениями об общем количестве пользователей, которое может существенно превышать количество тех, кто работает с продуктами «Телеком Дубль» (вполне вероятно, это свидетельствует об использовании продуктов компании-конкурента); данными о расходах на телекоммуникационные услуги, указанных в налоговых декларациях (появляется возможность сравнить их с расходами согласно биллингу компании); и общей информацией о доходах и расходах корпоративного клиента (позволяет спрогнозировать рост и сокращение расходов на предоставляемые услуги в будущем и предложить новые продукты).

Со временем в «Телеком Дубль» стало все больше ощущаться негативное влияние недостатков в управлении данными на выполнение стратегических задач. В частности, много неудобств доставляла практика размещения данных в изолированных хранилищах отдельных приложений – «бункерах» (silo), которыми не могут пользоваться другие приложения. В предисловии мы отметили отсутствие у руководства и сотрудников компании каких-то серьезных знаний в области управления данными. Ситуация начала меняться с появлением нового вице-президента, ответственного за работу с клиентами. Он уже обладал достаточным опытом в области современных методов управления информационными активами, который приобрел на предыдущих местах работы, в том числе и в зарубежных корпорациях. По инициативе нового вице-президента была создана экспертная рабочая группа по управлению данными, объединившая заинтересованных представителей как со стороны бизнеса, так и со стороны блока ИТ. Группе предстояло проанализировать вопросы создания стратегии работы с данными и программы управления данными (включая программу руководства данными).

Эксперты рабочей группы решили не приступать сразу к подготовке стратегии и программы, а провести ряд предварительных мероприятий. Их целью стало изучение и ослабление факторов, мешающих организации эффективно управлять данными, включая семь «смертных грехов», о которых мы говорили в главе 6. Такой подход основывался на собственном обобщенном представлении экспертной группы о состоянии практики управления данными в «Телеком Дубль» и мнении привлеченных внешних консультантов.

Для определения необходимых подготовительных мероприятий и последовательности их проведения был инициирован проект по разработке концепции создания корпоративной системы управления данными. В его рамках сотрудники компании (прежде всего эксперты рабочей группы) совместно с внешними консультантами проделали ключевые первоначальные шаги (более подробно их содержание рассматривается в главе 16 «Организационные аспекты управления данными»):

- провели оценку текущего состояния;
- выработали понимание возможностей по улучшению с целью разработки дорожной карты реализации программы управления данными;
- сформировали предложения по инициированию программы управления организационными изменениями.

В концепции (итоговом документе проекта) были приведены результаты выполненных шагов, включая обоснование необходимости стратегии работы с данными и программы управления данными, а также план проведения работ по их подготовке. После утверждения концепции высшим руководством рабочая группа приступила к реализации намеченных

мероприятий. Спустя некоторое время были разработаны и утверждены первые версии стратегии работы с данными и программы управления данными.

Стратегия работы с данными предусматривала порядок ее регулярного обновления. Напомним, что по мере роста зрелости практики управления данными в организациях фокус их стратегии в этой области меняется (см. рис. 6.7). Первая версия стратегии «Телеком Дубль» была сконцентрирована на повышении эффективности и результативности.

В следующих главах мы перейдем к обсуждению реализации компанией «Телеком Дубль» программы управления данными по ее отдельным направлениям.

Литература к главе 10

- Gilad U., Ashdown J., Lakshmanan V., Eryurek E., Kibunguchy-Grant A. Data Governance: The Definitive Guide: People, Processes, and Tools to Operationalize Data Trustworthiness. O'Reilly Media Inc., 2021.
- Marr B. Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things. Kogan Page, 2017.
- Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.
- Seiner R. Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success. Technics Publications, 2014.
- Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.
- Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярук. – М.: РАНХиГС, 2022.

Глава 11. Планирование и проектирование данных

В главе 9 «Управление данными: принципы и структуры» говорилось о том, что деятельность по непосредственному управлению жизненным циклом данных сосредоточена на трех направлениях, соответствующих основным фазам жизненного цикла (см. рис. 9.4):

1. планировании и проектировании данных;
2. обеспечении доступности и обслуживании данных;
3. практическом использовании данных и расширении возможностей их применения для достижения целей организации.

Сравнивая цепи поставок продукции с цепочками поставок данных, мы уже отмечали, что управление и тем и другим начинается с процессов планирования и проектирования (см. главу 7). В текущей главе мы обсудим это важное направление и такие вопросы, как:

- роль концепции архитектуры предприятия в планировании и проектировании различных сторон деятельности организации;
- основополагающее место архитектуры данных в управлении данными;
- цели моделирования данных и связанные с ним артефакты.

11.1. Архитектура данных

Чтобы подчеркнуть важность архитектурной практики в управлении данными, обратим внимание на следующие два момента.

1. В рамочной модели управления корпоративной информацией (рис. 7.7) частные политики, процессы, процедуры и стандарты, поддерживающие политику работы с информацией, опираются на информационную архитектуру, формируемую в соответствии с дата-центричным подходом (рис. 6.2).
2. В своей книге, посвященной стратегии работы с данными (на которую мы уже ссылались в главе 6), Айкен и Харбор в качестве одного из трех важнейших элементов эффективной работы с данными

(наряду с информационной грамотностью и цепочками поставок данных) выделяют стандартизацию информационных активов. Как мы увидим, в основе такой стандартизации лежит построение рациональной архитектуры данных.

11.1.1. Определение области знаний «Архитектура данных»

Архитектурой принято называть как искусство и науку о проектировании и строительстве, так и результаты этой деятельности – т.е. сами строения. В общем смысле архитектурой называют упорядоченную компоновку составляющих элементов, предполагающую оптимизацию функциональности, производительности, технологичности, стоимости и эстетичности создаваемой конструкции или системы.

Термин «архитектура» также применяется для описания отдельных аспектов проектирования информационных систем. Стандарт ISO/IEC/IEEE 42010:2011 Systems and software engineering. Architecture description («Системная и программная инженерия. Описание архитектуры») определяет архитектуру как «основные понятия или свойства системы в окружающей среде, воплощенной в ее элементах, отношениях и конкретных принципах ее проекта и развития»³⁸⁶.

На практике под термином «архитектура» в зависимости от контекста может подразумеваться описание текущего состояния систем, компонентов совокупностей систем, проектирование систем (как дисциплина и практика), планируемый проект системы или совокупности систем (будущее состояние или предполагаемая архитектура), артефакты, характеризующие систему (архитектурная документация), или команда, выполняющая работу по проектированию (архитекторы или архитектурная команда).

Архитектура данных образует фундамент управления данными. Поскольку большинство организаций располагают объемами данных, которые не могут быть осмыслены отдельными сотрудниками, возникает потребность в их представлении на разных уровнях абстракции таким образом, чтобы они были понятны, и руководство могло принимать относительно этих данных решения.

Архитектуру данных можно рассматривать с нескольких точек зрения, которые позволяют определить ее важнейшие составляющие.

- Выходные результаты архитектуры данных, такие как модели, определения и описания потоков данных на различных уровнях, т.е. все то, что принято называть артефактами архитектуры данных.
- Работы по формированию, развертыванию и внедрению целевых решений в области архитектуры данных.
- Организационное поведение в рамках работ в области архитектуры данных; формы сотрудничества, образы мышления и навыки, распределенные по различным ролям.

Архитектура данных организации описывается с помощью целостного комплекса проектных документов различной степени абстракции, включая стандарты, определяющие порядок сбора, хранения, упорядочения, использования и удаления данных. Она также делится на описания всех хранилищ данных и описания всех маршрутов их перемещения по информационным системам организации.

К артефактам архитектуры данных относятся спецификации, используемые для описания текущего состояния, определения требований к данным, порядка интеграции данных и контроля информационных активов в соответствии с действующей стратегией работы с данными. Наиболее детализированный архитектурный проектный документ в области данных – оформленная надлежащим образом корпоративная модель данных (включающая наименования элементов данных, подробные определения данных и метаданных, концептуальные и логические сущности и связи между ними, а также бизнес-правила). Наряду с другими документами в состав проектной документации входят и физические модели данных как продукты области моделирования и проектирования.

Архитектура данных наиболее полезна в тех случаях, когда она в полном объеме обеспечивает потребности на корпоративном уровне. Единая корпоративная архитектура данных позволяет последовательно и согласованно осуществлять стандартизацию и интеграцию данных в масштабах организации³⁸⁷.

³⁸⁶ См. ГОСТ Р 57100-2016/ISO/IEC/IEEE 42010:2011.

³⁸⁷ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

11.1.2. Цели и бизнес-драйверы

Основная цель архитектуры данных – служить мостом между бизнес-стратегией и ее технологической реализацией.

Будучи частью архитектуры предприятия (см. следующий раздел), архитектура данных должна:

- стратегически подготавливать организации к быстрому развитию продуктов, услуг и данных с целью полного использования бизнес-возможностей, которые открываются с появлением новых технологий;
- переводить бизнес-потребности на язык требований к данным и системам, чтобы бизнес-процессы не испытывали дефицита в необходимой информации;
- обеспечивать управление сложным процессом предоставления данных и информации в масштабах предприятия;
- способствовать повышению согласованности между бизнес– и ИТ-процессами;
- служить средством гибкого проведения изменений и преобразований.

Именно эти бизнес-драйверы определяют меру ценности архитектуры данных³⁸⁸.

11.1.3. Архитектура предприятия

Архитектура информационных систем обычно рассматривается в рамках более широкой дисциплины – архитектуры предприятия (Enterprise Architecture, EA)^{389 390}, которая охватывает архитектуры нескольких предметных областей (доменов), включая данные, бизнес-приложения и технологии.

Отлаженные практики управления архитектурой помогают организации четко понимать текущее состояние своих информационных систем, проводить изменения, направленные на переход в желаемое состояние, обеспечивать соблюдение нормативно-правовых требований, повышать эффективность и производительность работы. Эффективное управление данными и использование их и систем хранения – одна из общих целей разделов концепции архитектуры предприятия³⁹¹.

Архитектура данных строится в контексте других предметных областей архитектуры, включая бизнес, приложения и технологическую инфраструктуру. Таблица 11.1 содержит их сравнительное описание. Архитекторы, занимающиеся различными доменами, должны совместно определять направления и требования к разработке, согласовывая их между собой, поскольку каждый домен влияет и накладывает ограничения на другие.

11.1.4. Модель Захмана

Базовыми моделями, используемыми для разработки широкого спектра родственных архитектур, являются архитектурные рамочные структуры. Они представляют собой своего рода общую «архитектуру архитектуры» и служат способом осмысления и понимания архитектурных построений³⁹².

388 Там же.

389 В данном случае под «предприятием» понимается одна или несколько организаций, разделяющих определенную миссию, цели и задачи для получения результата в виде продукции или услуги. См., например, ГОСТ Р ИСО 15704-2008 «Промышленные автоматизированные системы. Требования к стандартным архитектурам и методологиям предприятия». Таким образом, в качестве предприятия может рассматриваться целая корпорация или ее подразделение, государственное учреждение и коммерческая фирма или, например, несколько фирм с общими владельцами.

390 В русском переводе DAMA-DMBOK2 применительно к терминам, начинающимся со слова enterprise в качестве определения (например, enterprise architecture или enterprise data model), это слово чаще всего переводится как «корпоративный» – «корпоративная архитектура», «корпоративная архитектура данных» и т. п. (широко распространенный вариант перевода в подобных случаях). Такой перевод облегчает восприятие текста в тех местах, где указанные термины применяются совместно с термином organization («организация»).

391 Там же.

392 Там же.

Людам, не связанным с архитектурой и не разбирающимся во всех различиях между архитектурными уровнями и аспектами, легко в них запутаться, поэтому бывает сложно понять, чем занимаются архитекторы. Одна из причин, по которой архитектурные рамочные структуры крайне полезны, состоит в том, что они помогают неспециалистам получить представление о связях и соотношениях между отдельными компонентами.

Таблица 11.1.

Архитектурные домены*

Домен	Бизнес-архитектура предприятия	Архитектура данных предприятия	Архитектура приложений предприятия	Технологическая архитектура предприятия
Назначение	Выявление путей создания предприятием ценности для потребителей и других заинтересованных лиц	Описание того, как данные должны быть организованы и как должно осуществляться управление данными	Описание структуры и функциональности используемых предприятием приложений	Описание технологической инфраструктуры, необходимой для обеспечения функционирования систем и предоставления с их помощью ценности
Элементы	Бизнес-модели, процессы, возможности, сервисы, события, стратегии, словарь	Модели данных, определения данных, спецификации отображения данных (data mapping), потоки данных, интерфейсы прикладного программирования (application programming interface, API) для работы со структурированными данными	Информационные системы, поддерживающие бизнес; пакеты прикладного ПО; базы данных	Технические платформы, сети, средства обеспечения безопасности и интеграции данных
Зависимости	Устанавливает требования для других доменов	Обеспечивает управление данными, создаваемыми и требуемыми согласно бизнес-архитектуре	Обеспечивает обработку данных в соответствии с бизнес-требованиями	Предоставление ресурсов для размещения решений, определенных архитектурой приложений
Роли	Бизнес-архитекторы и аналитики, распорядители бизнес-данных	Архитекторы данных и специалисты по разработке моделей данных, распорядители данных	Архитекторы приложений	Архитекторы инфраструктуры

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Самая известная рамочная структура архитектуры предприятия была разработана Джоном Захманом³⁹³ в 1980-х годах. Захман обратил внимание на то, что к созданию зданий, самолетов, предприятий, цепочек создания стоимости, проектов или систем имеют отношение различные группы людей, у каждой из которых есть собственная точка зрения на архитектуру создаваемого объекта. Эту концепцию он применил к требованиям для различных типов и уровней архитектуры предприятия.

Модель Захмана представляет собой матрицу 6×6, охватывающую полный набор моделей, требуемых для описания предприятия, и связей между ними. Архитектурная рамочная структура Захмана не описывает, как именно создавать входящие в нее модели. Она показывает, что эти модели должны быть созданы (табл. 11.2).

Столбцы матрицы отражают обсуждаемые вопросы (что, как, где, кто, когда и зачем), а строки – преобразования в процессе материализации (выявление – identification, определение – definition, представление, спецификация – specification, конфигурация – configuration, реализация – instantiation). В ячейках на пересечении строк и столбцов отражены уникальные типы артефактов архитектуры данных.

³⁹³ Джон Захман (John Zachman) – американский специалист по информационным технологиям для бизнеса, разработавший описываемую модель в рамках концепции планирования бизнес-систем IBM, одним из авторов которой является.

Таблица 11.2.

Упрощенное представление модели Захмана*

	Что	Как	Где	Кто	Когда
Руководство	Идентификация объектов	Идентификация процессов	Идентификация местоположений	Идентификация обязанностей	Идентификация ролей
Бизнес-менеджмент	Определение объектов	Определение процессов	Определение местоположений	Определение обязанностей	Определение ролей
Архитекторы	Представление объектов	Представление процессов	Представление местоположений	Представление обязанностей	Представление ролей
Инженеры	Спецификация объектов	Спецификация процессов	Спецификация местоположений	Спецификация обязанностей	Спецификация ролей
Технические специалисты	Конфигурация объектов	Конфигурация процессов	Конфигурация местоположений	Конфигурация обязанностей	Конфигурация ролей
Предприятие	Реализация объектов	Реализация процессов	Реализация местоположений	Реализация обязанностей	Реализация ролей
Наборы объектов	Потоки объектов	Потоки процессов	Расположение	Распределение обязанностей	Распределение ролей

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Обсуждаемые аспекты представляют собой вопросы, которые могут быть заданы относительно какой-либо сущности. Применительно к архитектуре предприятия столбцы матрицы можно интерпретировать следующим образом:

- **что** (столбец объектов): сущности (объекты), используемые для построения архитектуры;
- **как** (столбец процессов): проводимые работы;
- **где** (столбец местоположений): местоположения бизнес-структур и технологических структур;
- **кто** (столбец обязанностей): роли и организационные системы;
- **когда** (столбец привязки по времени): сроки, интервалы, события, циклы, расписания;

- **зачем** (столбец мотивации): цели, стратегии и средства.

Процесс материализации состоит из шагов, необходимых для перевода абстрактной идеи в конкретный образец (реализация). Эти шаги отражены в строках матрицы, обозначенных с помощью названий соответствующих ролей: планировщик, владелец, проектировщик, разработчик, внедренный пользователь. Каждой из перечисленных ролей соответствует отличная от других перспектива процесса в целом, а также собственный круг решаемых проблем. Эта специфика и показана в строках. Например, каждая перспектива выражает различное отношение к столбцу «Что» (предметы или данные).

- **Перспектива руководства** (бизнес-контекст). Перечни составляющих бизнеса, определяющие содержание моделей идентификации.

- **Перспектива бизнес-менеджмента** (бизнес-концепции). Уточнение бизнес-менеджерами (как владельцами) связей между бизнес-концепциями и отражение их в моделях определения.

- **Перспектива архитектора** (бизнес-логика). Логические системные модели, детализирующие системные требования, и проектные решения без учета ограничений, отраженные архитекторами (как проектировщиками) в моделях представления.

- **Перспектива инженера** (физический уровень). Физические модели, оптимизирующие проектные решения с целью их реализации для конкретных применений с учетом ограничений по используемым технологиям, человеческим ресурсам, стоимости и срокам. Определяются инженерами (как разработчиками) в моделях спецификации.

- **Перспектива технического специалиста** (сборка). Чисто технический, без учета контекста, взгляд на то, каким образом отдельные компоненты должны быть собраны и функционировать. Отражается техническими специалистами (как внедренцами) в конфигурационных моделях.

- **Перспектива пользователя** (реализация). Реальные функционирующие объекты, с которыми работают сотрудники (как пользователи). Эта перспектива моделей не предусматривает.

Как уже отмечалось, каждой ячейке, определяемой в результате пересечения строки и столбца, в модели Захмана соответствует уникальный тип разрабатываемого артефакта. Каждый такой артефакт описывает, каким образом соответствующая перспектива отвечает на вопросы, обсуждаемые в процессе создания архитектуры.

11.1.5. Основные артефакты архитектуры данных

По мере поступления данных в организацию через каналы связи или интерфейсы обеспечивается их защита и интеграция; они сохраняются, регистрируются, каталогизируются, распространяются, включаются в отчеты, анализируются и предоставляются заинтересованным лицам. Попутно данные могут подвергаться верификации, улучшению, связыванию, сертификации, агрегированию, анонимизации и использованию в целях аналитики вплоть до момента их архивации или удаления. Следовательно, описания корпоративной архитектуры данных должны включать как корпоративные модели данных (с указанием структуры и спецификаций данных), так и описания потоков данных.

Корпоративная модель данных и описание потоков данных должны быть хорошо согласованы. При этом и модель, и потоки данных отражаются в трех состояниях – текущем, целевом (архитектурная перспектива) и переходном (проектная перспектива)³⁹⁴.

Корпоративная модель данных

Корпоративная модель данных (Enterprise Data Model, EDM) представляет собой целостную, не зависящую от технических средств реализации концептуальную или логическую модель данных, отражающую единый согласованный взгляд на данные в масштабах всей организации. Этот термин обычно используется для обозначения высокоуровневой упрощенной модели данных, но уровень абстракции может быть различным в зависимости от целей ее представления. EDM содержит данные о ключевых сущностях предприятия (на уровне бизнес-концепций) и связях между ними, критически важные руководящие бизнес-правила и некоторые ключевые атрибуты. EDM закладывает основу для всех проектов в области данных или связанных с данными. Модели данных уровня отдельных проектов должны создаваться на основе EDM (см. главу 6 – разделы 6.3 и 6.4 и главу 7 – раздел 7.3). Данная модель подлежит

³⁹⁴ Там же.

обязательной проверке всеми заинтересованными сторонами для обеспечения согласованного мнения о том, что в ней зафиксировано правильное представление об организации.

Организация, осознавшая потребность в EDM, должна определить, сколько времени и усилий она готова посвятить ее построению и ведению. EDM могут создаваться с различными уровнями детализации, а потому нужно изначально определиться с имеющимися ресурсами и, исходя из этого, спланировать объем работ по подготовке первоначального содержания модели. Со временем по мере необходимости можно расширять объемы и прорабатывать дополнительные детали собираемых данных, требующихся для оптимальной работы организации, что, как правило, и делается. Самые успешные EDM выстраиваются поэтапно, итерационно и послойно.

Рисунок 11.1 показывает, как связаны модели различных типов и как концептуальные модели могут быть привязаны к физическим моделям данных приложений. На рисунке отражены следующие уровни представления:

- концептуальная общая модель данных, предоставляющая обзор всех предметных областей организации;
- представления сущностей и связей по каждой предметной области;
- детализированные, с частично описанными атрибутами логические представления тех же предметных областей;
- логические (Logical Data Model, LDM) и физические (Physical Data Model, PDM) модели на уровне отдельных приложений или проектов.

Все уровни в совокупности составляют корпоративную модель данных. Структура связей позволяет проследить сущность с верхнего уровня до нижнего и между моделями на одном уровне.

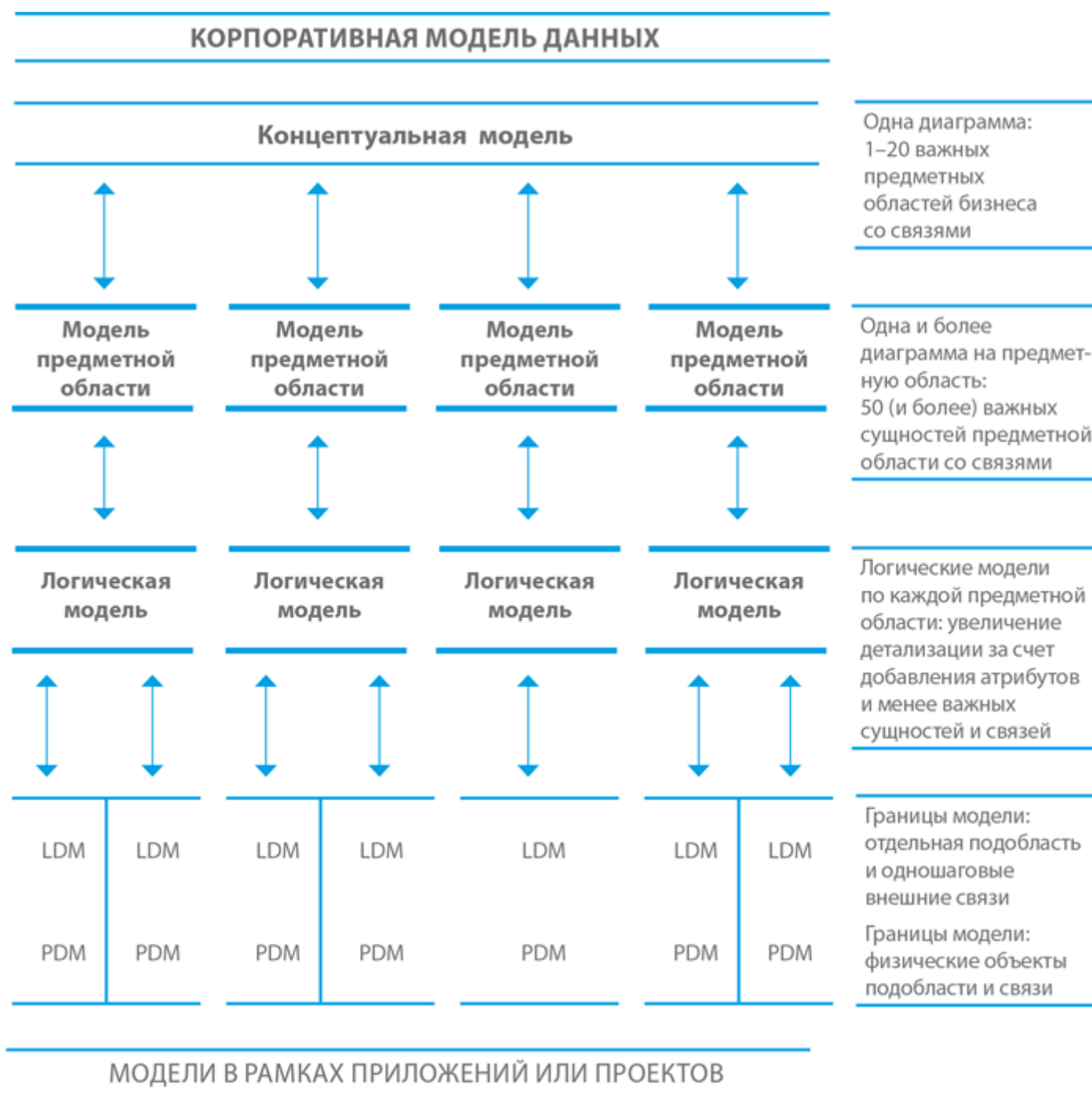


Рис. 11.1.

Корпоративная модель данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Описание потоков данных

В главе 8 мы подчеркивали важность понятий «поток данных» («цепочка данных») и «происхождение данных». Описание потоков данных содержит требования и основное рабочее описание (master blueprint) организации хранения и обработки данных по всем базам данных, приложениям, платформам и сетям. Потоки данных отражают их перемещение с целью использования в бизнес-процессах, на отдельных рабочих местах, сотрудниками с определенными бизнес-ролями, а также отдельными техническими компонентами.

Потоки данных являются одним из способов документального оформления происхождения данных. Они фиксируют маршруты прохождения данных через бизнес-процессы и системы. Описанный от начала до конца поток данных показывает, где данные возникают, где хранятся и используются, а также все преобразования данных в процессе их движения как внутри, так и между различными процессами и системами. Анализ происхождения помогает объяснить состояние данных в каждой точке потока.

Потоки данных отображают и документируют взаимосвязи данных:

- с приложениями, используемыми в рамках бизнес-процесса;
- хранилищами или базами данных в среде функционирования;
- сегментами сети (полезно для описания мер безопасности);
- бизнес-ролями, показывая, какие роли отвечают за создание, чтение, обновление, удаление данных;
- местами, в которых происходят изменения данных.

Потоки данных могут документироваться с разной степенью детализации – до уровня предметной области, сущности или даже атрибута. Системы могут быть представлены сегментами сети, платформами, наборами часто используемых приложений или отдельными серверами. Для схематического представления потоков данных могут использоваться матрицы (рис. 11.2) или диаграммы потоков данных (рис. 11.3).

11.1.6. Две точки зрения на архитектуру данных

В главе 8 в качестве наиболее поздних этапов эволюции развития управления данными были выделены этапы 2 и 3 (табл. 8.1). Второй этап ориентирован прежде всего на качество данных, а третий – на инновации на основе данных. В связи с этим создание архитектуры данных (и архитектуры предприятия) сопряжено с необходимостью учета сложного комплекса вопросов, обусловленных двумя основными точками зрения на архитектурные решения.

		БИЗНЕС-ПРОЦЕССЫ						
		Проектирование продукта	Маркетинг и продажи	Подготовка производства	Управление заказами	Производство	Логистика	Выставление счетов
ОСНОВНЫЕ СУЩНОСТИ	Продукт	●	●	●	●	●		
	Комплектуемый элемент	●	●	●	●	●		
	Производственное предприятие	●		●	●	●	●	
	Клиент		●		●	●	●	●
	Товарная позиция		●		●		●	●
	Структура сборки	●		●		●		
	Заказ на поставку		●		●	●	●	●
	Заказ на производство			●	●			
	Индивидуальный продукт					●	●	●
	Поставка						●	
	Счет клиента		●					●
		●	●	●	●	●	●	●

● Создание
 ● Чтение/использование

Рис. 11.2.

Поток данных, представленный в виде матрицы*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

1. **Ориентированная на качество.** Основное внимание направлено на совершенствование деятельности в рамках бизнес-цикла и цикла разработки в области ИТ. Без должного управления

архитектурой качество решений ухудшается. Системы со временем будут чрезмерно усложняться и терять гибкость, что создаст дополнительные риски для организации. Неконтролируемое распространение и копирование данных, запутанные взаимосвязи делают организации менее эффективными и снижают доверие к данным.

2. Ориентированная на инновации. Основное внимание направлено на трансформацию бизнеса и ИТ в свете новых перспектив. Продвигать инновации за счет внедрения прорывных технологий и методов использования данных — еще одна задача современного корпоративного архитектора.

К двум этим позициям нужно подходить дифференцированно.

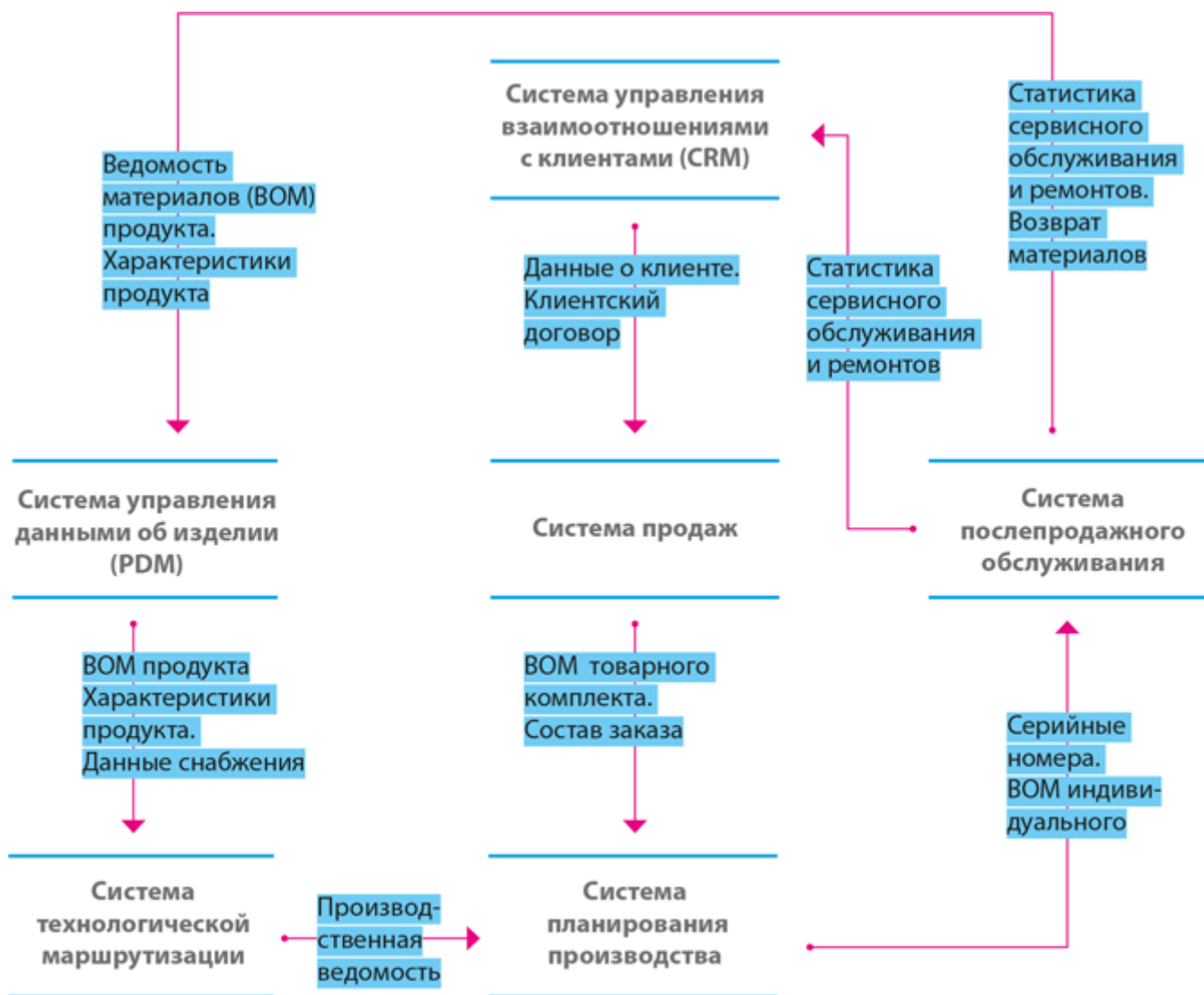


Рис. 11.3.

Пример диаграммы потоков данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Подход, ориентированный на качество, укладывается в традиционное представление о работе по проектированию архитектуры, предусматривающее ее поэтапное совершенствование. Архитектурные

задачи распределяются по проектам, в которых принимают участие архитекторы. В любом случае специалист, как правило, не теряет из виду цельность архитектуры и руководствуется долгосрочными установками, связанными с руководством данными, стандартизацией и структурированной разработкой. Такой подход в большей степени способствует выработке дата-центричной архитектуры, о которой мы говорили в главах 6 и 7.

При подходе, ориентированном на инновации, может рассматриваться краткосрочная перспектива, а также предполагается использование еще не апробированных схем ведения бизнеса и передовых технологий. Такая направленность часто требует, чтобы архитекторы вступали в контакт с теми людьми внутри организации, которые обычно не входят в круг общения профессионалов в сфере ИТ (например, с представителями подразделений, отвечающих за разработку новых продуктов или бизнес-моделей)³⁹⁵.

11.1.7. Контекстная диаграмма области знаний и уровни зрелости функции «Архитектура данных»

Контекстная диаграмма области знаний «Архитектура данных» представлена на рисунке 11.4.

Работая в команде по поддержке корпоративной архитектуры (если в организации внедрена функция поддержки архитектуры предприятия) или в команде по поддержке архитектуры данных, архитекторы данных отвечают за разработку дорожной карты, управление требованиями к корпоративным данным в рамках проектов и интеграцию с общей архитектурой предприятия. Успех зависит от определения и соблюдения архитектурных стандартов, а также от создания и поддержания полезных и пригодных для использования архитектурных артефактов. Соблюдение дисциплины в архитектурной практике может повысить ее эффективность и качество за счет создания повторно используемых и расширяемых решений (см. главу 6 – разделы 6.3 и 6.4 и главу 7 – раздел 7.3).

Важный аспект проводимых работ – предоставление экспертам в отдельных предметных областях организации возможности сотрудничать в решении архитектурных задач. При наличии соответствующего времени и ресурсов это приведет к тому, что оптимальные архитектурные решения будут определяться в рамках обычных проектных циклов, а не путем приложения экстраординарных усилий при первоначальном проектировании.

Цель состоит в том, чтобы иметь запас созданных архитектурных решений, готовых к объединению и выполнению будущих бизнес-задач.

395 Там же.

АРХИТЕКТУРА ДАННЫХ

Определение: определение потребностей организации в данных (безотносительно к структуре), а также разработка и сопровождение основных рабочих описаний решений по их обеспечению. Использование основных рабочих описаний в качестве руководящих материалов при осуществлении интеграции данных и контроля информационных активов, а также при согласовании инвестиций в области данных с бизнес-стратегией

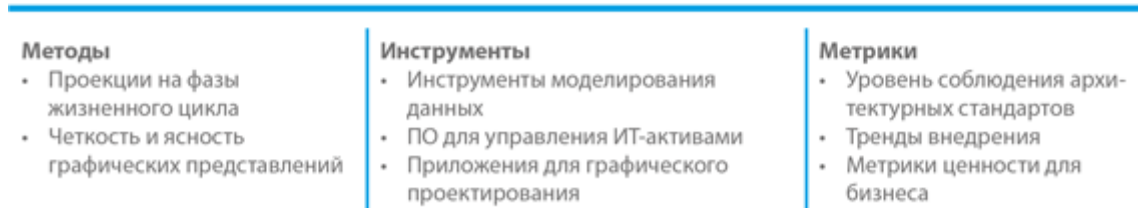
Цели:

1. Определение требований к хранению и обработке данных
2. Разработка структур и планов, направленных на обеспечение текущих и долгосрочных потребностей организации в данных
3. Обеспечение стратегической готовности организации к быстрому развитию своих продуктов, услуг и данных с целью получения преимуществ от использования возможностей, заложенных в новейших технологиях

БИЗНЕС-ДРАЙВЕРЫ



ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 11.4.

Контекстная диаграмма. Архитектура данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Один из способов устранения препятствий для совместной работы – создание в организации единой

стандартной архитектурной рамочной структуры, позволяющей использовать общий словарь, а также шаблоны, облегчающие кросс-функциональную интеграцию данных.

На рисунке 11.5 представлены обобщенные характеристики уровней зрелости функции «Архитектура данных».



Рис. 11.5.

Обобщенные характеристики уровней зрелости функции «Архитектура данных»*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

11.1.8. Влияние на ценность данных

В каждой организации реализована некоторая архитектура данных, но для многих – это плохо представляемый комплекс проектных решений, про которые известно только то, что они состоят из некоторого количества физических хранилищ данных, систем, процессов, сервисов и интерфейсов. Данные поступают в организацию, циркулируют в ней и выходят через эти компоненты.

В таких организациях обычно мало что из архитектурного ландшафта задокументировано, поэтому никто не может с достаточной уверенностью утверждать, каким образом данные поддерживают бизнес. Это также означает, что изменения в информационном ландшафте могут оказаться слишком рискованными.

Во введении мы говорили о том, что четвертая промышленная революция характеризуется дизруптивным (ломающим привычные представления) воздействием на компании. Последние, чтобы остаться на плаву, вынуждены в ускоренном темпе трансформировать себя. У организаций, которые не способны внедрять инновации, быстро реагировать на меняющиеся рыночные условия, будет самое трудное и неопределенное будущее.

Архитектура данных в целом направлена на развитие способностей организации эффективно работать с данными и, следовательно, на повышение готовности к изменениям ее деятельности.

Архитекторы данных создают и поддерживают знания организации о данных и системах, через которые эти данные распространяются. Такие знания позволяют управлять данными как активом и повышать получаемую от них выгоду за счет выявления возможностей по их применению, а также снижения издержек и рисков. Поэтому одним из артефактов архитектуры данных (наряду с корпоративной моделью данных, описанием потоков данных и другими) могут быть описания цепочек ценности данных, которые мы обсуждали в главе 7.

Архитектуру стремятся сделать такой, чтобы она приносила организации ценность. Ценность же достигается за счет оптимизации требуемых ресурсов, операционной и проектной эффективности, а также расширения возможностей организации по использованию данных (см. главу 5). Для этого необходимы качественное проектирование и планирование, а также способность обеспечить эффективную реализацию проектов и планов³⁹⁶. Кроме того, повышение ценности данных обеспечивается за счет расширения сотрудничества между различными функциональными направлениями организации – налаживания отношений, выявления областей для сокращения дублирования и затрат. Это позволит привлечь к участию больше сотрудников и повысить заинтересованность в улучшении процессов управления данными. Таким образом, архитектура данных станет важным элементом поддержки непрерывного совершенствования.

11.2. Моделирование и проектирование данных

Моделирование данных – критически важный компонент управления данными. Оно требует от организации выяснения и документирования того, как данные соотносятся друг с другом в рамках общей картины. В то же время моделирование само по себе заключается в разработке решений в отношении компоновки данных и их связи. Модели данных отражают и одновременно улучшают понимание организацией информационных активов, которыми она располагает и оперирует^{397, 398}.

11.2.1. Определение области знаний «Моделирование и проектирование данных»

Моделирование данных – это процесс последовательного выявления, анализа и формулирования основных требований к данным с последующим их представлением и распространением в точно определенной форме.

³⁹⁶ Там же.

³⁹⁷ Там же.

³⁹⁸ Kent W. Data and Reality: A Timeless Perspective on Perceiving and Managing Information in Our Imprecise World, 3rd Edition. Technics Publications, 2012.

Модель – это представление чего-либо, что уже существует, или примерный образец того, что предстоит создать. Модель может содержать одну или несколько диаграмм. В каждой диаграмме используются стандартные символы, обеспечивающие понимание ее смыслового содержания. Примерами широко распространенных моделей являются карты, схемы организационных структур, чертежи зданий.

Модели данных – важная форма метаданных. Они содержат полезные для потребителей данных сведения. Значительная часть этих сведений, выявленных в процессе моделирования, необходима другим функциям управления данными. Например, определения, требующиеся для руководства данными, или информация, относящаяся к происхождению данных и используемая при ведении хранилищ данных, а также в бизнес-аналитике.

Модель данных либо описывает данные организации так, как они понимаются на текущий момент, либо отражает то состояние данных, в котором организация хотела бы их видеть. Она содержит набор символов с текстовыми метками, предназначенными для визуального представления требований к данным, в том виде, в котором их сообщили специалисту по моделированию. При этом количество элементов описываемой области данных может варьироваться от небольшого (если рассматривается отдельный проект) до весьма внушительного (если рассматривается организация).

Модель данных – форма документирования требований к данным и определений данных. Получаемые в результате процесса моделирования документально оформленные модели – главное средство коммуникации, обеспечивающее передачу требований к данным от сферы бизнеса в блок ИТ, а также (в рамках блока ИТ) от аналитиков, специалистов по моделированию и архитекторов, взаимодействующих с ними, проектировщикам и разработчикам баз данных.

11.2.2. Цели и бизнес-драйверы

Главная цель моделирования данных – подтвердить и документально зафиксировать понимание различных аспектов организации данных, которое обеспечит создание приложений, наиболее точно соответствующих текущим и будущим потребностям бизнеса, а также заложить фундамент для успешной реализации широкомасштабных инициатив, таких как программы управления основными данными и руководства данными.

Модели данных имеют критическое значение для эффективного управления данными, поскольку они:

- определяют единую общую терминологию во всем, что касается данных;
- собирают и документируют точные знания (метаданные) о данных и информационных системах организации;
- служат основным средством коммуникации в процессе реализации проектов;
- являются отправной точкой при настройке, интеграции или замене приложений³⁹⁹.

11.2.3. Разновидности моделей данных

Существует множество различных типов моделей данных. Наиболее распространенные из них: реляционная, многомерная, объектно-ориентированная, на основе фактов, хронологическая и NoSQL⁴⁰⁰. Разработчики информационных систем используют модели в зависимости от потребностей, а также особенностей моделируемых данных и систем организации, для которой разрабатывается модель. Различные типы моделей данных визуально представляют данные с помощью различных соглашений.

Модели также различаются в зависимости от уровня абстракции объектов, которые они отображают: концептуальная с высоким уровнем абстракции; логическая со средним уровнем абстракции и физическая, которая отображает конкретную систему или экземпляр данных. При разработке информационных систем концептуальное и логическое моделирование данных относят к категории работ по планированию и

³⁹⁹ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁰⁰ От англ. not only SQL – «не только SQL». SQL (Structured Query Language) – структурированный язык запросов, используемый в традиционных системах управления базами данных на основе реляционной модели.

анализу требований, а физическое моделирование данных – к проектным работам⁴⁰¹⁴⁰².

11.2.4. Строительные блоки моделей данных

Вне зависимости от типа моделей, в большинстве из них выделяются одни и те же компоненты – «строительные блоки»: сущности, связи, атрибуты и области значений атрибута. Приведенные здесь определения и примеры помогут составить представление о том, как работают модели данных^{403, 404}.

Сущность

В общем смысле – вне контекста моделирования данных – под сущностью понимается предмет, существующий отдельно от других предметов. В рамках моделирования данных сущность – предмет, о котором организация собирает информацию. Иногда сущности уподобляют «существительным» организации. Действительно, сущность можно рассматривать как ответ на один из фундаментальных вопросов (кто, что, где, когда, почему и как) или сочетание таких ответов. В таблице 11.3 приведены определения и примеры общеупотребительных категорий сущностей⁴⁰⁵.

401 Поэтому рассматриваемая область знаний называется не просто «Моделирование данных», а «Моделирование и проектирование данных».

402 Там же.

403 Там же.

404 Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

405 DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Таблица 11.3.

Общепотребительные категории сущностей

Категория	Определение	Примеры
Кто	Физическое лицо или организация, представляющие интерес. Тот, <i>кто</i> важен для бизнеса. Категория «кто» часто представляет однородные группы или роли, такие как клиент или поставщик. При этом любое лицо или организация могут быть включены в различные группы и выступать в различных ролях	Работодатель, пациент, игрок, подозреваемый, клиент, поставщик, студент, пассажир, конкурент, автор
Что	Продукт, услуга, товар или иной предмет интереса предприятия. То, <i>что</i> важно для бизнеса. Обычно используется для определения категорий производимой продукции или услуг, оказываемых организацией. Здесь крайне важно четко определять атрибуты категорий, типов и т. п.	Продукт, услуга, поставка, товар, курс, саундтрек, фотография, книга
Когда	Календарные даты, сроки или периоды, интересующие организацию. <i>Когда</i> именно действует то, что важно для бизнеса	Время, дата, месяц, квартал, год, расписание, семестр, срок, время отправления
Где	Места локализации интересов организации, включая фактические физические, почтовые и электронные адреса. <i>Где</i> ведется бизнес	Почтовый адрес, пункт выдачи, URL веб-сайта, IP-адрес
Почему	События или транзакции, представляющие интерес для организации и держащие бизнес на плаву. <i>Почему и для чего</i> делается то, что делается	Заказ, возврат, претензия, жалоба, депозит, отзыв, запрос, замена, рекламация
Как	Документация, относящаяся к событиям, интересующим организацию. Документы служат подтверждением факта того или иного события (например, оформления заказа или его исполнения). Как мы определяем, имело ли место то или иное событие	Счет-фактура, договор, контракт, заказ на покупку, квитанция об уплате штрафа за нарушение, товарная накладная, подтверждение сделки
Измерение	Итоги, суммы и т. п. Сводные и контрольные данные по всем остальным категориям	Продажи, количество единиц товара, платежи, баланс

Связь

Связь – это отношение между сущностями. Связи фиксируют информацию о взаимодействиях между концептуальными сущностями, детализированные характеристики взаимодействия между логическими сущностями и взаимные ограничения при взаимодействии физических сущностей. Связи на диаграммах моделей данных принято отображать линиями.

Важная характеристика связи – ее мощность. Мощность связи между двумя сущностями определяет, сколько экземпляров одной сущности и сколько экземпляров другой могут быть связаны друг с другом. Например, в организации может работать один или несколько сотрудников.

Мощность отображается специальными символами («вилками») на обоих концах линии связи. Допустимые значения мощности – ноль, один или много («много» означает «больше чем один»). Возможны произвольные сочетания трех этих значений на противоположных концах связи.

На рисунке 11.6 показаны различные соотношения значений мощности связи на примере

реляционной модели данных⁴⁰⁶. В организации работает один или несколько сотрудников. Сотрудник может содержать ноль, одного или нескольких иждивенцев. Но сотрудник занимает одну и только одну должность в течение определенного периода времени. Мощность связи – это способ фиксации правил и предположений, связанных с данными. Если данные показывают, что сотрудник выполняет более одной работы в течение установленного периода времени, то в них имеется ошибка или в организации допускаются отклонения от правила.

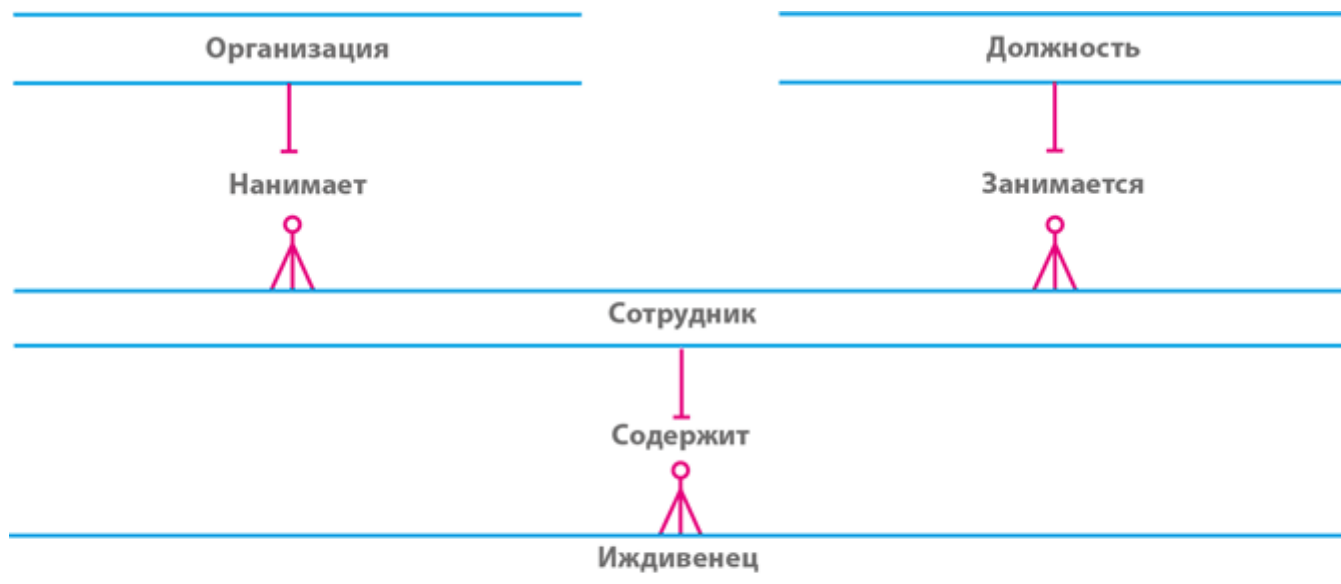


Рис. 11.6.

Соотношения значений мощности связи в модели данных*

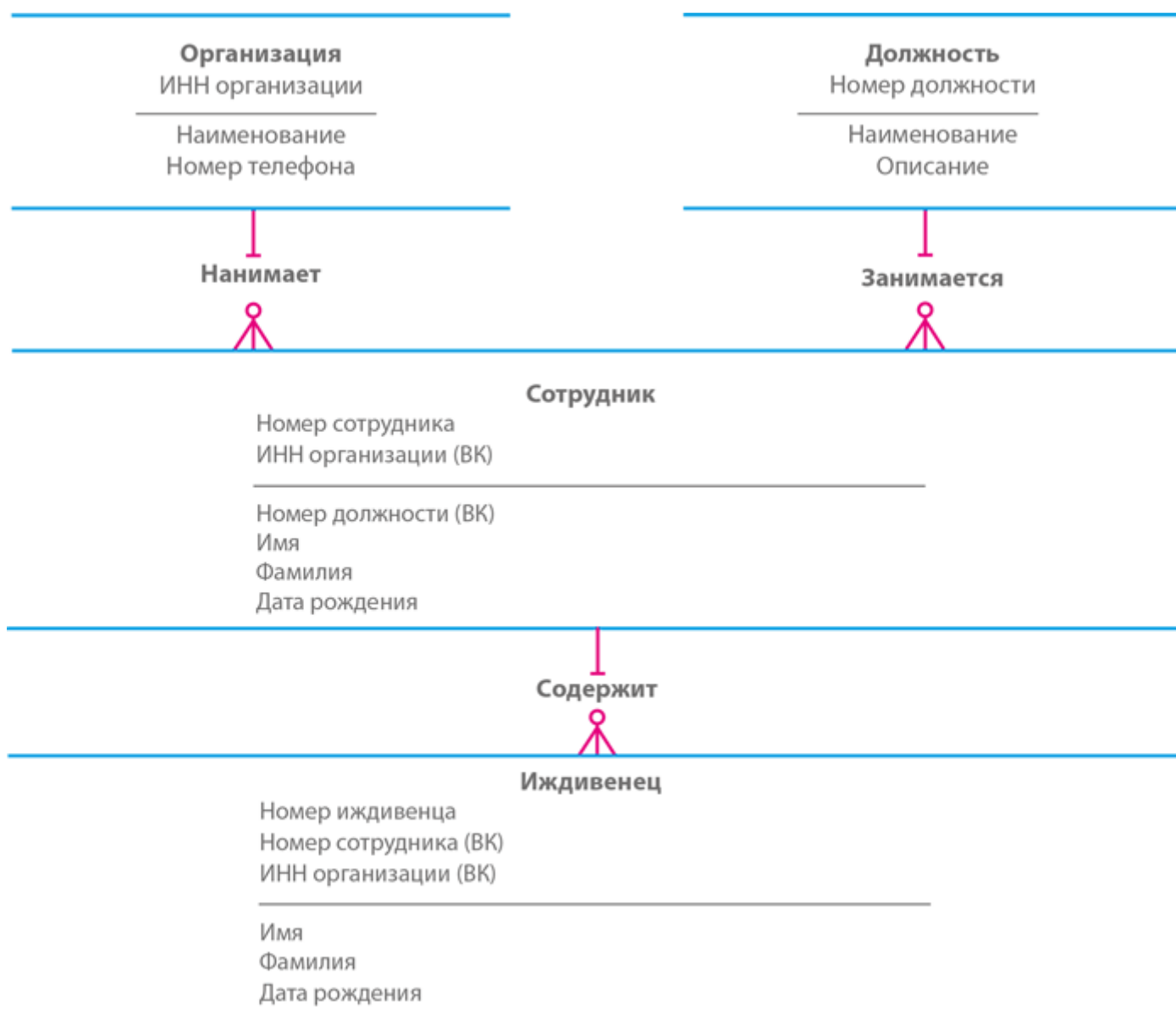
* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

Атрибут

Атрибут – это характеристика сущности, позволяющая ее идентифицировать, описать или измерить. На физическом уровне атрибуту сущности может соответствовать столбец, поле, тег или узел в таблице, представлении, документе, графе или файле.

На рисунке 11.7 представлены сущности с описывающими их атрибутами (на примере реляционной модели данных). Сущность «Организация» имеет атрибуты «ИНН организации», «Наименование» и «Номер телефона». Сущность «Сотрудник» имеет атрибуты «Номер сотрудника», «Имя», «Фамилия» и «Дата рождения». Сущности «Иждивенец» и «Должность» имеют атрибуты, отражающие их основные характеристики.

⁴⁰⁶ Реляционная модель данных (от англ. relation – «отношение», «зависимость», «связь») представляет собой совокупность взаимосвязанных двумерных таблиц (в терминах теории реляционных баз данных – «отношений»), каждая из которых содержит сведения об объектах определенного типа (товарах, клиентах и т. п.).

**Рис. 11.7.****Атрибуты сущностей и ключи***

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

На представленной диаграмме атрибуты «ИИН организации», «Номер сотрудника» и «Номер должности» являются первичными ключами соответствующих сущностей. Ключом называют атрибут или набор атрибутов, уникальным образом определяющий экземпляр сущности. Поскольку в общем случае вариантов ключей (так называемых потенциальных ключей) может быть несколько, то один из них выбирается в качестве фактического уникального идентификатора экземпляра – первичного ключа. В сущности «Сотрудник» и «Иждивенец» для организации связей с другими сущностями (расположенными на диаграмме над ними) добавлены так называемые внешние ключи. Атрибут (или набор атрибутов) сущности, который является внешним ключом, предназначен для хранения значения первичного ключа другой сущности. У каждого экземпляра сущностей «Сотрудник» и «Иждивенец» значение внешнего ключа должно совпадать со значением первичного ключа одного из экземпляров соответствующих связанных сущностей.

Домен

Отметим, что в моделировании данных доменом обычно называется исчерпывающим образом

описанный набор, диапазон или множество значений, которые могут быть присвоены атрибуту. В свою очередь, определение домена – одно из средств стандартизации характеристик атрибутов. Например, домен «Дата», включающий все допустимые значения календарных дат, может задаваться для любого атрибута датировки в логической модели и для любых столбцов/полей дат в физической модели данных, таких как:

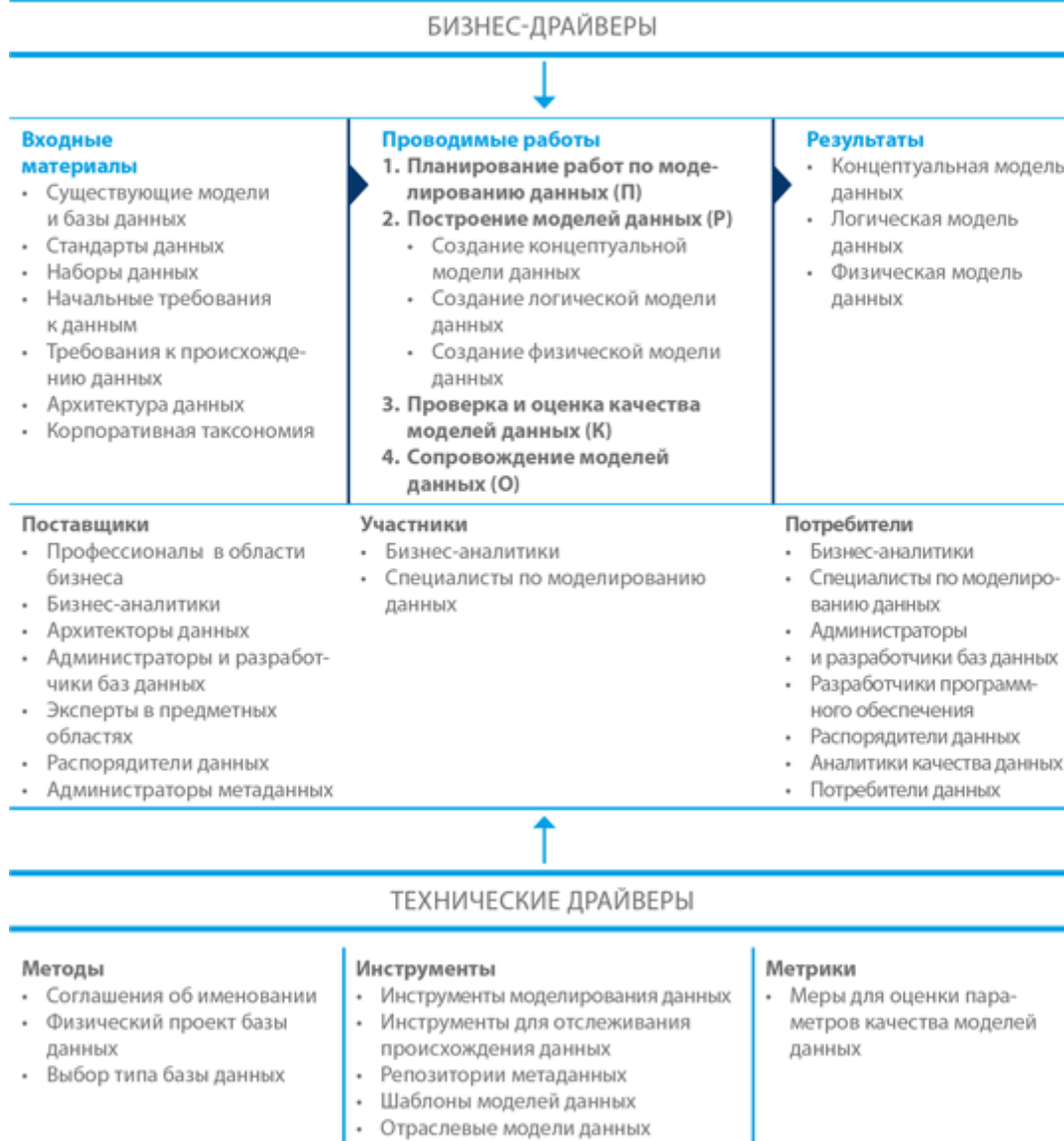
- дата_приема_на_работу;
- дата_поступления_заказа;
- дата_рекламации;
- дата_начала_занятий.

Домены важны для понимания качества данных. Все значения, входящие в домен, являются допустимыми значениями. Те, которые выходят за его границы, – недопустимы. Домен для атрибута «дата_приема_на_работу» может быть определен просто как действительные даты. Согласно этому правилу, он, например, не включает 30 февраля любого года.

МОДЕЛИРОВАНИЕ И ПРОЕКТИРОВАНИЕ ДАННЫХ

Определение: моделирование данных — процесс выявления, анализа и формулирования требований к данным с последующим их представлением и распространением в точно определенной форме, называемой моделью данных. Этот процесс носит итерационный характер и может включать разработку концептуальной, логической и физической моделей

Цели: подтвердить и документально зафиксировать понимание различных аспектов организации данных, которое обеспечит создание приложений, наиболее точно соответствующих текущим и будущим потребностям бизнеса, а также заложить фундамент для успешной реализации широкомасштабных инициатив, таких как программы управления основными данными и руководства данными



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 11.8.

Контекстная диаграмма. Моделирование и проектирование данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

11.2.5. Контекстная диаграмма области знаний и уровни зрелости функции «Моделирование и

проектирование данных»

Контекстная диаграмма области знаний «Моделирование и проектирование данных» представлена на рисунке 11.8.

Аналитики данных, разработчики моделей и баз данных выступают в роли посредников между потребителями информации (теми, кто определяет нужды бизнеса в данных) и производителями данных (теми, кто фиксирует данные в пригодной для использования форме). Профессионалы в области данных должны обеспечивать искомый баланс при учете требований к данным от потребителей информации и требований к приложениям от производителей данных.

Но и профессионалы, работающие в области данных, также должны обеспечивать баланс – причем с учетом краткосрочных и долгосрочных интересов бизнеса. Потребителям информации нужны актуальные данные для выполнения своих обязанностей по текущему управлению бизнесом и реализации возможностей. Команды проектов по созданию систем должны укладываться в заданные временные и бюджетные рамки. Они должны учитывать интересы всех заинтересованных сторон, обеспечивая размещение данных организации в безопасных и надежных хранилищах, защищенных системами резервного копирования и обеспечивающих совместный доступ к данным и их повторному использованию, а также корректность, актуальность, релевантность и максимальное удобство использования данных с точки зрения пользователей. Именно поэтому модели и проектные решения по организации базы данных должны быть разумно сбалансированы таким образом, чтобы учитывать как краткосрочные, так и долгосрочные нужды организации.

На рисунке 11.9 представлены обобщенные характеристики уровней зрелости функции «Моделирование и проектирование данных».

11.2.6. Влияние на ценность данных

Вполне осязаемые результаты правильного моделирования данных: снижение затрат на поддержку, расширение возможности повторного использования моделей при проведении в жизнь будущих инициатив, минимизация затрат на создание новых приложений.

Подтверждение и документирование понимания различных аспектов организации данных и перспектив в рамках моделирования данных способствует более эффективной деятельности по следующим направлениям.

Повторное использование артефактов моделирования из репозитория проходит циклы обратной связи с надежным контролем управления версиями. При реализации проектов применяются в качестве эталонных повторно используемые решения.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Практики моделирования и проектирования данных измеряются и контролируются согласно общеорганизационным стандартам, при этом все бизнес-функции отчитываются совету по руководству данными. Существует наполненный репозиторий артефактов моделирования, и организовано их согласованное повторное использование всеми бизнес-функциями.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Определены общеорганизационные политики и стандарты, устанавливающие стратегическую линию в отношении применения практик моделирования и проектирования данных по всем бизнес-функциям. Подразделения устанавливают свои локальные практики на основе этих стандартов и должны использовать инструменты и методологии, позволяющие обмениваться моделями в рамках всей организации.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Практики моделирования и проектирования данных в рамках конкретных бизнес-функций, подразделений или отдельных проектов успешно внедрены, задокументированы и повторяются. Хотя на этих уровнях инструменты и методологии стандартизованы, нет никаких свидетельств их координирования на более высоком уровне, что приводит к непоследовательным подходам или дублированию усилий в рамках всей организации.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В подразделениях или при взаимодействии между ними отсутствует какая-либо стандартизация в отношении практики моделирования данных.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 11.9.

Обобщенные характеристики уровней зрелости функции «Моделирование и проектирование данных»*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

- **Формализация.** Модель данных документирует краткое и четкое определение структур данных и связей между ними. Она позволяет оценивать, как влияют на данные реализованные бизнес-правила (как для текущих, так и для будущих целевых состояний). Формальное определение вводит строго соблюдаемую структуру данных, что снижает вероятность нарушений при обеспечении доступа к данным и их ведении. Иллюстрируя структуры данных и связи между их элементами, модель данных упрощает их практическое использование.

- **Определение области применения.** Модель данных помогает объяснить границы контекста данных,

внедрения приобретенного программного обеспечения и области охвата проектов, инициатив и существующих систем.

- **Сохранение/документирование знаний.** Модель данных может сохранять корпоративную память о какой-либо системе или проекте, фиксируя знания в четко определенной форме. Она служит документацией для будущих проектов в качестве версии «как есть».

Модели данных помогают лучше понимать различные аспекты организации или бизнеса, механизмы работы приложений или последствия изменений существующей структуры данных. Таким образом, модель данных становится многократно используемой картой, помогающей профессионалам в области бизнеса, руководителям проектов, аналитикам, специалистам по моделированию и разработчикам лучше понимать структуру данных в контексте среды окружения. Так же как картографы изучают и документируют географический ландшафт, помогая другим осуществлять навигацию, специалисты по моделированию данных помогают другим понять информационный ландшафт⁴⁰⁷.



ПРАКТИЧЕСКИЙ ПРИМЕР

С этого блока нашего сквозного примера мы начинаем обсуждение реализации программы управления данными компании «Телеком Дубль».

В рамках мероприятий по планированию и проектированию данных в компании началась работа по созданию корпоративной модели данных и описанию потоков данных. Также в подразделениях началось обсуждение основных цепочек ценности данных.

Специалисты «Телеком Дубль» приступили к переходу на дата-центричную архитектуру, о которой мы говорили в главе 6 (см. рис. 6.2). Далее, в главе 13 мы рассмотрим основные домены (предметные области) телекоммуникационной компании (см. раздел 13.8). По каждому из этих доменов, и в первую очередь по клиентскому, разрабатываются модели данных, ориентированные на использование всеми системами.

При внедрении новых приложений компания не расширяет ИТ-ландшафт для дополнения сведений о клиенте новыми данными во всех смежных системах (CDI, CRM, ERP), а централизованно обновляет модель данных клиентского домена, что позволяет бесшовно для систем-потребителей получать новую информацию о клиентах.

Литература к главе 11

- Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synch in the Post-Big Data Era. Technics Publication, 2017.
- Bernard S. An Introduction to Holistic Enterprise Architecture: Fourth Edition. AuthorHouse, 2020.
- Fox R. Controlling the Chaos: A Functional Framework for Enterprise Architecture and Governance; First Edition. Technics Publications, 2018.
- Hoberman S. Data Modeling Made Simple: A Practical Guide for Business and IT Professionals. Second Edition. Technics Publications, 2009.
- Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.
- Strengtholt P. Data Management at Scale: Best Practices for Enterprise Architecture; 1st Edition. O'Reilly Media, Inc., 2010.
- Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

Глава 12. Обеспечение доступности и обслуживание данных: основы

⁴⁰⁷ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

В этой главе (и двух следующих) будет рассмотрена очень важная и довольно многочисленная группа областей знаний (функций) по управлению данными. Эти области (как и описанные в предыдущей главе области архитектуры и моделирования данных), относятся к непосредственному управлению жизненным циклом данных. Они отвечают за укрупненную фазу жизненного цикла, идущую вслед за фазой планирования и проектирования – обеспечение доступности и обслуживание данных (см. рис. 9.4). В рассматриваемую группу входят:

- хранение и операции с данными;
- интеграция и интероперабельность данных;
- управление справочными и основными данными;
- ведение хранилищ данных;
- хранение больших данных;
- управление документами и контентом⁴⁰⁸.

В главе 7 мы сравнивали процессы цепочки поставок данных с процессами цепи поставок продукции (SCOR-моделью). Если взглянуть на перечисленные области в этом контексте, то их можно соотнести со следующими процессами:

- снабжать,
- делать,
- доставлять,
- возвращать.

В упрощенной модели цепочки поставок данных, предлагаемой Дагласом Лейни (см. рис. 7.3), представленные области распространяются преимущественно на этапы сбора и администрирования.

Данная глава посвящена первым трем функциональным компонентам из рассматриваемой группы. Они закладывают основу для обеспечения доступности и обслуживания данных.

12.1. Хранение и операции с данными

В качестве первого звена в цепочках ценности или поставок всегда выступают процессы сбора (см. главу 7). Однако перед тем как ввести их в действие, необходимо решить вопросы организации хранения и обслуживания собираемых материальных ресурсов или данных. Хранение и выполнение операций с данными – это то, что прежде всего представляют себе многие люди, когда слышат об управлении данными. Именно с функций хранения и администрирования началось развитие этой дисциплины (см. табл. 8.1).

12.1.1. Определение области знаний «Хранение и операции с данными»

Область «Хранение и операции с данными» включает проектирование и реализацию решений для хранения, а также сопровождение хранимых данных с целью получения от них максимальной выгоды на протяжении всего их жизненного цикла. Работы в этой области ведутся по двум основным направлениям.

● **Сопровождение баз данных**⁴⁰⁹. Объединяет работы, относящиеся к жизненному циклу данных, включая первоначальную реализацию рабочей среды базы данных (database environment), получение данных, а также их резервное копирование и удаление. Сюда же относится обеспечение оптимальной производительности (мониторинг и настройка – критически важные элементы сопровождения).

⁴⁰⁸ В DAMA-DMBOK2 эта функция отнесена к фазе жизненного цикла, связанной с использованием и расширением возможностей применения данных (см. главу 8). Однако логичнее отнести ее к предыдущей фазе – обеспечения доступности и обслуживания, как это сделано в книге Себастьян-Коулман*, которая представляет собой краткий обзор DAMA-DMBOK2 для руководителей.

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition, 2018.

⁴⁰⁹ Под базой данных (БД) понимается любая совокупность хранимых данных, вне зависимости от их структуры или содержания*.

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

● **Технологическая поддержка баз данных** включает определение технических требований, соответствующих информационным потребностям организации, определение технической архитектуры, развертывание и администрирование технологических решений, а также разрешение проблемных вопросов, связанных с технологиями.

12.1.2. Цели и бизнес-драйверы

Цели хранения и операций с данными включают:

- управление доступностью данных на протяжении всего их жизненного цикла;
- обеспечение целостности информационных активов;
- управление эффективностью проведения информационных транзакций.

В процессе операционной деятельности организации постоянно используют собственные информационные системы. С учетом этого обстоятельства хранение и операции с данными являются жизненно важными аспектами деятельности организаций. Таким образом, обеспечение непрерывности бизнеса – главный драйвер усилий в рассматриваемой области управления данными. Если база данных оказывается недоступной, текущая операционная деятельность организации осуществляется с задержками или останавливается. Надежная инфраструктура хранения данных, обеспечивающая проведение операций, позволяет свести к минимуму риск подобных сбоев⁴¹⁰.

12.1.3. Архитектуры и модели организации баз данных

С архитектурной точки зрения базы данных подразделяются на два типа: централизованные и распределенные. Централизованная система управления базами данных (СУБД) управляет базами данных, которые реализованы в одном месте, в то время как распределенная система управляет множеством баз данных, реализованных во множестве систем.

Распределенные системы можно разделить на два класса по степени автономности входящих в них компонентов: федеративные (автономные компоненты) и не федеративные (неавтономные компоненты).

Архитектуры федеративных систем баз данных различаются в зависимости от уровней интеграции с локальными базами данных и объема предлагаемых услуг. В целом федеративные СУБД можно разделить на слабо связанные и сильно связанные⁴¹¹.

В настоящее время широкое распространение получила виртуализация (также называемая облачными вычислениями, cloud computing), которая позволяет оказывать услуги по проведению вычислений, использованию программного обеспечения, предоставлению доступа к данным и их хранению таким образом, что конечному пользователю не требуются знания о физическом местонахождении и конфигурации систем, обеспечивающих предоставление этих услуг.

Ниже кратко описаны некоторые методы реализации баз данных в облаке.

● **Образ виртуальной машины.** Облачные платформы предоставляют пользователям возможность арендовать экземпляры виртуальных машин и использовать их для работы со своими базами данных. Пользователи могут либо загружать на них собственные образы машины с развернутой базой данных, либо использовать предлагаемые провайдерами готовые образы машин с предустановленными и настроенными СУБД.

● **База данных как услуга.** Некоторые облачные платформы предлагают возможность использования базы данных как услуги (Database-as-a-Service, DaaS) без запуска экземпляра виртуальной машины. В такой конфигурации владельцы приложения вовсе избавлены от необходимости устанавливать и поддерживать базу данных. Провайдер услуги DaaS сам устанавливает и поддерживает базу данных, а владельцы приложения пользуются ею за абонентскую плату.

● **Управляемый облачный хостинг базы данных.** При таком варианте база данных не предлагается в качестве услуги; вместо этого провайдер облачного сервиса размещает ее у себя в облаке и осуществляет управление базой данных по поручению и в интересах собственника приложения⁴¹².

410 Там же.

411 Там же.

412 Там же.

Что касается моделей организации баз данных, то в предыдущей главе мы уже отмечали, что наиболее распространенными из них являются: реляционная, многомерная, объектно-ориентированная, на основе фактов, хронологическая и NoSQL.

12.1.4. Администраторы баз данных

Ключевую роль в каждом из направлений деятельности в области хранения и операций с данными играют администраторы баз данных (АБД). АБД – наиболее устоявшаяся и общепринятая профессиональная роль в сфере управления данными.

Наряду с АБД в некоторых организациях предусмотрены роли администраторов сетевых систем хранения данных, которые специализируются на сопровождении сетевых хранилищ, рассматриваемых отдельно от остальных приложений или структур, обеспечивающих хранение данных.

Функция хранения и выполнения операций с данными – это технически сложная работа, выполняемая АБД и администраторами сетевых систем хранения для обеспечения доступности и высокой производительности баз данных, а также сохранения их целостности.

Администрирование баз данных иногда рассматривается как единая монолитная функция, но АБД выступают в разных ролях. Они могут сопровождать среды эксплуатации баз данных, выполнять работы по разработке или поддерживать конкретные приложения и процедуры. На профиль работы администратора баз данных влияет общая архитектура баз данных организации (централизованная, распределенная, федеративная; слабо или сильно связанная), а также то, какие модели в них применены (реляционная, объектно-ориентированная, NoSQL и т. п.). С появлением новых технологий АБД и администраторы сетевых систем хранения стали отвечать за создание виртуальных сред и управление ими (облачные вычисления). Поскольку среды хранения данных довольно сложны, администраторы баз данных ищут способы уменьшить сложность или, по крайней мере, управлять ею за счет автоматизации, возможностей повторного использования и применения стандартов и передовых практик.

Хотя может показаться, что администраторы баз данных не связаны с функцией руководства данными, их знания технической среды необходимы для реализации руководящих установок в отношении данных, связанных с такими аспектами, как контроль доступа или конфиденциальность и безопасность данных. Опытные администраторы баз данных также играют важную роль в расширении возможностей организации по внедрению и использованию новых технологий^{413, 414}.

12.1.5. Контекстная диаграмма области знаний и уровни зрелости функции «Хранение и операции с данными»

Контекстная диаграмма области знаний «Хранение и операции с данными» представлена на рисунке 12.1.

АБД совместно с архитекторами данных вносят вклад в обеспечение доступности и обслуживание данных, работая по следующим направлениям:

- определение требований к хранению данных;
- определение требований к доступу к данным;
- разработка конкретных реализаций (экземпляров) баз данных;
- управление физической средой хранения;
- загрузка данных;
- репликация данных – ведение дублирующих друг друга баз данных (реплик);
- отслеживание шаблонов использования (типичных схем распределения нагрузки на базы данных);
- планирование обеспечения непрерывности бизнеса;
- управление резервным копированием и восстановлением;
- управление производительностью и доступностью баз данных;
- управление альтернативными средами баз данных (например, для разработки и тестирования);

413 Там же.

414 Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition, 2018.

- управление миграцией данных (переносом из одних хранилищ в другие);
- отслеживание и учет информационных активов;
- аудит и проверка корректности данных.

На рисунке 12.2 представлены обобщенные характеристики уровней зрелости функции «Хранение и операции с данными».



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 12.1.

Контекстная диаграмма. Хранение и операции с данными*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)



Рис. 12.2.

Обобщенные характеристики уровней зрелости функции «Хранение и операции с данными»*

* Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

12.1.6. Влияние на ценность данных

Непрерывность информационного обслуживания – важнейший фактор, позволяющий организации достигать своих целей и поддерживать удовлетворенность заинтересованных сторон. Основательный и продуманный подход к сопровождению и технологической поддержке баз данных, включая планирование мероприятий по обеспечению непрерывности бизнеса в случае аварий или чрезвычайных ситуаций, позволяет избежать сбоев в обслуживании (либо в результате превышения объема данных над имеющейся емкостью хранилищ, либо в результате катастрофического инцидента).

Переход к виртуализированным средствам хранения и обработки данных обеспечивает целый ряд вполне ощутимых преимуществ.

- Возможность достижения значительной экономии средств, поскольку использованием баз данных можно управлять более эффективно, а сопровождение осуществляется более централизованно.
- Подготовка к аварийному восстановлению упрощается, когда все информационные ресурсы, подлежащие включению в резервную копию, находятся в одном месте.
- По аналогичным соображениям упрощается создание многоуровневой архитектуры интеграции с предоставлением централизованных информационных сервисов (более подробно см. в разделе 12.2).
- С точки зрения руководства данными централизация обеспечивает более эффективный мониторинг безопасности, производительности и соответствия нормативно-правовым требованиям.

12.2. Интеграция и интероперабельность данных

На рисунке 9.4, который, как было замечено, по сути дела, отражает модель управления цепочками поставок данных, отсутствует специально выделенная функциональная область, отвечающая за их сбор. Однако это не означает, что вопросы сбора данных оставлены без внимания. Просто они распределены между различными функциональными элементами представленной схемы. Одну из главных ролей в обеспечении сбора данных играет функция обеспечения интеграции и интероперабельности.

12.2.1. Определение области знаний «Интеграция и интероперабельность данных»

Интеграция и интероперабельность данных (Data Integration and Interoperability, ДИ) – область, которая описывает процессы, связанные с перемещением и консолидацией данных как внутри хранилищ, приложений и организаций, так и в рамках обеспечения их взаимодействия.

Интеграция позволяет объединять данные в согласованные физические или виртуальные формы. Под интероперабельностью данных подразумевается способность двух или более информационных систем или компонентов к обмену и использованию информации, полученной в результате обмена^{415, 416}.

12.2.2. Цели и бизнес-драйверы

Внедрение практик и решений в области интеграции и интероперабельности данных преследует следующие цели:

- своевременное предоставление требуемых данных потребителям (как пользователям, так и приложениям) в нужном им формате;
- физическая или виртуальная консолидация данных в хабах (концентраторах);
- снижение стоимости и сложности решений по управлению данными за счет разработки общих моделей и интерфейсов;
- выявление значимых событий (возможностей и угроз) и автоматический запуск процедур выдачи уведомлений и принятия мер;
- поддержка функций BI, аналитики, управления основными данными и обеспечение операционной

⁴¹⁵ ГОСТ Р 55062-2012. Информационные технологии (ИТ). Системы промышленной автоматизации и их интеграция. Интероперабельность. Основные положения.

⁴¹⁶ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

эффективности.

Основной драйвер деятельности в области интеграции и интероперабельности данных – потребность в управлении перемещением данных.

Поскольку в большинстве организаций имеются сотни, а то и тысячи всевозможных баз и хранилищ данных, управление процессами перемещения данных между местами хранения внутри организации и обмена данными с другими организациями становится одной из главных сфер ответственности любой ИТ-службы. Без надлежащего управления процесс перемещения данных быстро исчерпает все их ресурсы и возможности, лишив при этом необходимой поддержки традиционные приложения и области управления данными.

Повсеместный переход организаций на использование покупного прикладного ПО вместо разработки собственного усилил потребность в обеспечении интеграции и интероперабельности на корпоративном уровне. Каждое коммерческое приложение добавляет собственный набор хранилищ основных, транзакционных и отчетных данных. Все их приходится интегрировать с другими хранилищами данных, уже имеющимися в организации. Даже системы планирования ресурсов предприятия (ERP), обеспечивающие выполнение общих функций организации, практически никогда не охватывают всех необходимых хранилищ данных. Они также должны интегрировать свои данные с другими данными организации.

Еще один важнейший бизнес-драйвер интеграции – управление затратами на поддержку. Перемещение данных с использованием множества технологий, каждая из которых требует специфических навыков разработки и обслуживания, способно привести к непомерному росту стоимости поддержки. Внедрение стандартных инструментов позволяет сократить потребности в обслуживании и персонале, а также повысить эффективность поиска и устранения неполадок, что дает возможность перераспределить ресурсы сопровождения на решение других приоритетных задач организации.

Проведение работ в области интеграции и интероперабельности данных также помогает организации соблюдать действующие стандарты и регламенты обработки данных. Интеграционные системы корпоративного уровня позволяют повторно использовать коды, обеспечивающие соответствие требованиям нормативных документов, и упрощают проверку их соблюдения.

12.2.3. Основные аспекты обеспечения интеграции и интероперабельности данных

При создании решений в области интеграции и интероперабельности данных важно уделять внимание следующим аспектам^{417, 418}.

Корпоративный подход

При проектировании интеграционных решений следует придерживаться корпоративного (в масштабах всей организации) подхода, обеспечивающего возможность последующего расширения и масштабирования, но реализацию проводить итерационно, методом пошагового ввода новых решений в эксплуатацию. Центральное место в таком подходе занимает идея минимизации дублирования уже затраченных на интеграцию усилий.

Важным инструментом в реализации корпоративного подхода являются корпоративные сервисные шины (Enterprise Service Buses, ESB) – интеграционные решения, обеспечивающие синхронизацию данных в режиме, близком к реальному времени, между многими системами. Такие решения используют понятие хаба данных, предоставляющего каноническую модель для совместного использования данных организацией.

Каноническая модель данных – общая модель (используемая организацией или группой, отвечающей за обмен данными), стандартизирующая формат, в котором осуществляется распространение данных. Использование канонической модели ограничивает количество преобразований данных при обмене между системами или организациями. Каждой системе достаточно реализовать преобразование данных только в

⁴¹⁷ Там же.

⁴¹⁸ Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

каноническую модель (при передаче) или из нее (при приеме), вместо того чтобы разрабатывать отдельные средства преобразования для множества систем, с которыми осуществляется обмен. В средах, где обмениваются данными более ста прикладных систем, интеграционное решение на основе канонической модели – единственно возможное.

ESB – пример реализации подхода к построению интеграционных решений, основанного на слабом связывании. Она действует как сервис обмена данными между приложениями. При таком подходе получение ответов на запросы, обращенные к другой системе, не является обязательным условием продолжения работы первой системы, т. е. доступность каждой из слабо связанных систем не зависит от доступности другой системы. Слабое связывание может быть реализовано с использованием различных средств: например, посредством сервисов, интерфейсов прикладного программирования (API) или очередей сообщений.

Исследование данных

Исследование данных необходимо проводить перед проектированием интеграционных решений. Цель исследования – определение потенциальных источников данных, которые могут быть использованы при выполнении работ по интеграции. Оно должно выявить, где данные могут быть получены и где они должны интегрироваться.

В крупных организациях существуют сотни, если не тысячи, источников данных, которые могут быть полезны для различных подразделений. Во многих случаях эти источники предоставляют одни и те же данные, но каждый из них доступен только в рамках отдельных проектов.

Сервис-ориентированная архитектура

Наиболее зрелые корпоративные стратегии интеграции приложений используют концепцию сервис-ориентированной архитектуры (SOA), в которой функциональность по предоставлению или обновлению данных может быть представлена в виде точно определенных вызовов сервисов, используемых приложениями в процессе их взаимодействия. При таком подходе приложениям не нужно взаимодействовать друг с другом напрямую или знать что-либо о внутренней структуре и работе других приложений. SOA обеспечивает независимость приложений и возможность замены той или иной системы в организации без необходимости внесения существенных изменений в системы, которые с ней взаимодействуют.

Цель сервис-ориентированной архитектуры – организация строго определенного взаимодействия между отдельными независимыми программными модулями. Каждый модуль выполняет функции (часто говорят «предоставляет сервисы») в интересах других программных модулей или людей. Ключевой концептуальный момент SOA – предоставляемые сервисы независимы: сервис и приложение ничего не знают друг о друге. Сервис-ориентированная архитектура может быть реализована с помощью различных технологий, включая веб-сервисы и обмен сообщениями.

Сами сервисы обычно реализуются как API, доступные для вызова прикладным системам или пользователям (потребителям). Регистрационная запись точно определенного API описывает доступные опции, необходимые параметры запроса и выдаваемую в ответ на обращение информацию.

Примерами наиболее часто применяемых стандартов реализации являются:

- SOAP: простой протокол доступа к объектам (Simple Object Access Protocol) – протокол обмена структурированными сообщениями в распределенной вычислительной среде;
- RESTful API: набор архитектурных принципов построения сервис-ориентированных приложений. REST – сокр. от англ. Representational State Transfer (передача состояния представления). RESTful – прилагательное, употребляющееся по отношению к сервисам, которые соответствуют принципам REST;
- JMS: служба сообщений Java (Java Message Service) – стандарт обмена сообщениями между приложениями, выполненными на платформе Java;
- RMI: удаленный вызов методов (Remote Method Invocation) – программный интерфейс для вызова удаленных процедур на языке Java.

Модель публикации и подписки

Модель публикации и подписки (publish and subscribe) предусматривает наличие систем,

поставляющих данные («издателей»), и систем, получающих эти данные («подписчиков»). Системы, поставляющие данные, вносятся в каталог сервисов данных, а системы, которым эти данные требуются, должны подписываться на услуги провайдера. После публикации данные автоматически рассылаются подписчикам.

При наличии множества потребителей одних и тех же наборов данных или данных в одном и том же формате подготовка этих данных в централизованном порядке (с последующим открытием доступа к ним) позволяет обеспечивать использование потребителями согласованных наборов данных и их регулярное своевременное обновление.

Модель публикации и подписки идеально подходит для распространения данных среди всех заинтересованных сторон.

Извлечение, преобразование и загрузка

В основе любых решений в области интеграции и интероперабельности данных лежит процесс извлечения, преобразования и загрузки (Extract, Transform, Load; ETL). Вне зависимости от того, выполняются они физически или виртуально, в пакетном режиме или режиме реального времени, эти шаги непременно присутствуют при перемещении данных между приложениями и организациями.

Процесс преобразования переводит выбранные данные в структуру, совместимую с целевым хранилищем. Часто бывает так, что при этом нужно объединить фрагменты данных вместе (агрегирование) или, возможно, выполнить операции с данными, или провести вычисления, чтобы предоставить дополнительную информацию (обогащение). Границы между преобразованием, агрегированием и обогащением провести непросто, но все эти действия представляют собой добавление некоторой ценности к исходным данным. Это позволяет представлять потребителям данные в более полезной форме.

Задержка при обработке

В зависимости от требований по интеграции данных процедуры ETL могут выполняться в режиме периодической пакетной обработки или обработки по мере доступности новых или обновленных данных (в режиме реального времени или управляемой на основе событий – event driven). Обработка данных о текущих операциях обычно проводится в режиме реального времени или в режиме, близком к реальному времени (near real-time), а данных, требуемых для анализа и отчетности, – по графику, в пакетном режиме.

Обычное явление сегодня – потоковая обработка данных. Поточковые данные (streaming data) «вытекают» из компьютерных систем в непрерывном режиме по ходу событий (фиксируется такая информация, как сведения о покупках товаров или ценных бумаг, комментарии в социальных сетях или показания датчиков, отслеживающих различные характеристики). Однако реализация потоковой обработки сопряжена с серьезными затратами на аппаратное и программное обеспечение.

Задержка (latency) – это разница во времени между моментом, когда данные были сгенерированы в системе-источнике, и моментом, когда они стали доступны в целевой системе. Различные подходы к обработке данных определяют различную степень задержки.

Задержка может быть высокой (при пакетной обработке), низкой (при запуске процедур переноса на основе событий) или очень низкой (при использовании синхронизации в режиме реального времени или при потоковой обработке).

Оркестровка данных

Потоки данных в интеграционном решении должны быть спроектированы и документально оформлены. Оркестровка данных как раз и представляет собой описание потоков данных от «старта» до «финиша», включая промежуточные шаги, требуемые для выполнения преобразования и транзакции. Можно рассмотреть, например, такой набор действий, которые могут образовывать единую транзакцию: разместить заказ, произвести оплату, запросить доставку, отменить заказ, вернуть платеж, отменить доставку. Оркестровка пакетной интеграции данных должна также предоставлять сведения о частоте перемещения и преобразования данных. Отдельные задачи, с помощью которых реализуется пакетная интеграция, обычно описываются в планировщике, который и запускает их в указанное время, с указанной периодичностью или по наступлении заданного события. Расписание задач может включать множество взаимозависимых шагов.

Оркестровка интеграции данных в режиме реального времени, как правило, предусматривает запуск задач по событию – например, добавлению или обновлению данных. Такая оркестровка обычно сложнее, чем в пакетном режиме, и реализуется посредством применения многих инструментов.

Одна из главных задач оркестровки – обеспечить, чтобы каждое из отдельных действий, выполняемых в рамках потока, в случае какого-либо сбоя было завершено корректно и согласованно, а целостность данных во всех взаимодействующих системах сохранились.

Проверка качества данных

Сервис-ориентированный подход подразумевает внедрение элементов стандартизации, что облегчает деятельность по контролю и повышению качества данных. Это связано с тем, что все данные, проходящие через централизованные сервисы, могут быть проверены на соответствие правилам валидации, что позволяет обнаруживать, обрабатывать и сообщать об имеющихся ошибках.

В результате любые системы, подписанные на услуги по предоставлению данных, будут получать данные, уровень качества которых измерен и известен.

Таким образом, интеграционная архитектура – важный компонент повышения качества данных и может уменьшить необходимость инвестиций в применяемые для этой цели автономные инструменты.

В завершение обсуждения ключевых аспектов функции обеспечения интеграции и интероперабельности данных следует заметить, что она критически важна для ведения хранилищ данных и бизнес-аналитики, а также для управления справочными и основными данными, поскольку обе эти области управления данными сфокусированы на преобразовании и интеграции данных из систем-источников в консолидационных хабах, с последующей передачей консолидированных данных в целевые системы, которые предоставляют их потребителям. На рисунке 12.3 приведен пример представления целевой многоуровневой интеграционной архитектуры, спроектированной с учетом перечисленных выше аспектов.

Диаграммы подобного рода могут быть полезны при объяснении всем заинтересованным сторонам ключевого принципа развития интеграционных решений – устранение связей «точка-точка» за счет реализации более многоуровневой технологии, поддерживаемой ESB.

12.2.4. Контекстная диаграмма области знаний и уровни зрелости функции «Интеграция и интероперабельность данных»

Контекстная диаграмма области знаний «Интеграция и интероперабельность данных» представлена на рисунке 12.4.

Интеграция и интероперабельность данных зависит от других областей управления данными:

- **руководство данными** – в части определения правил преобразования данных и структуры сообщений;
- **архитектура данных** – в части разработки архитектуры интеграционных решений;
- **безопасность данных** – в части обеспечения соответствия интеграционных решений требованиям по безопасности данных, как постоянно хранимых (persistent), так и виртуальных (virtual), а также «данных в движении» (in motion), которые перемещаются между приложениями и организациями;
- **метаданные** – в части отслеживания такой информации, как техническое описание данных (постоянно хранимых, виртуальных и передаваемых), описание их значения для бизнеса, описание бизнес-правил преобразования данных, а также история операций и сведения о происхождении (lineage) данных;
- **хранение и операции с данными** – в части физической реализации решений по хранению данных;
- **моделирование и проектирование данных** – в части проектирования структур данных (постоянно хранимых, виртуальных, а также сообщений, которые перемещаются между приложениями и организациями).

На рисунке 12.5 представлены обобщенные характеристики уровней зрелости функции «Интеграция и интероперабельность данных».

ОПЕРАЦИОННЫЕ СИСТЕМЫ

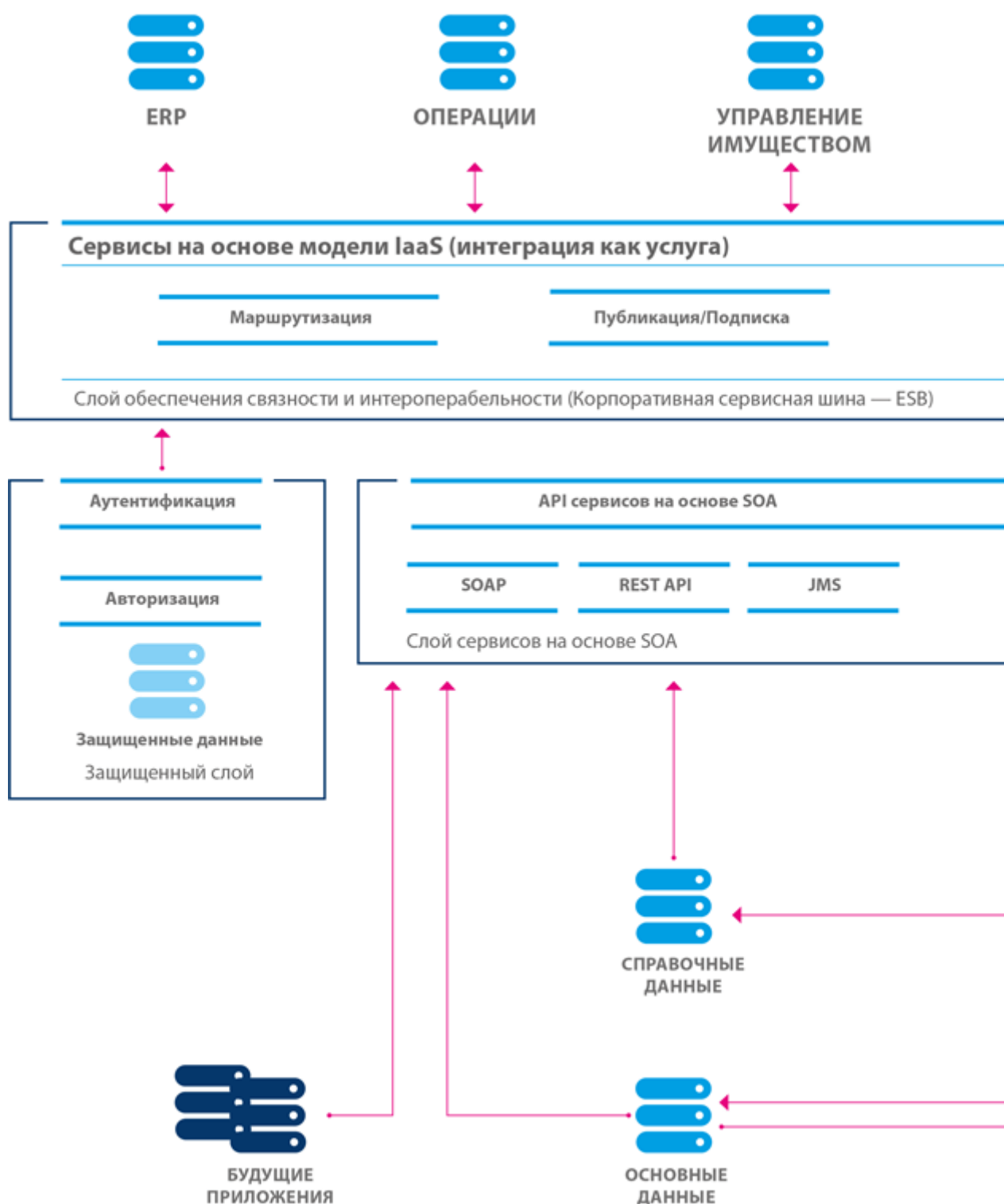
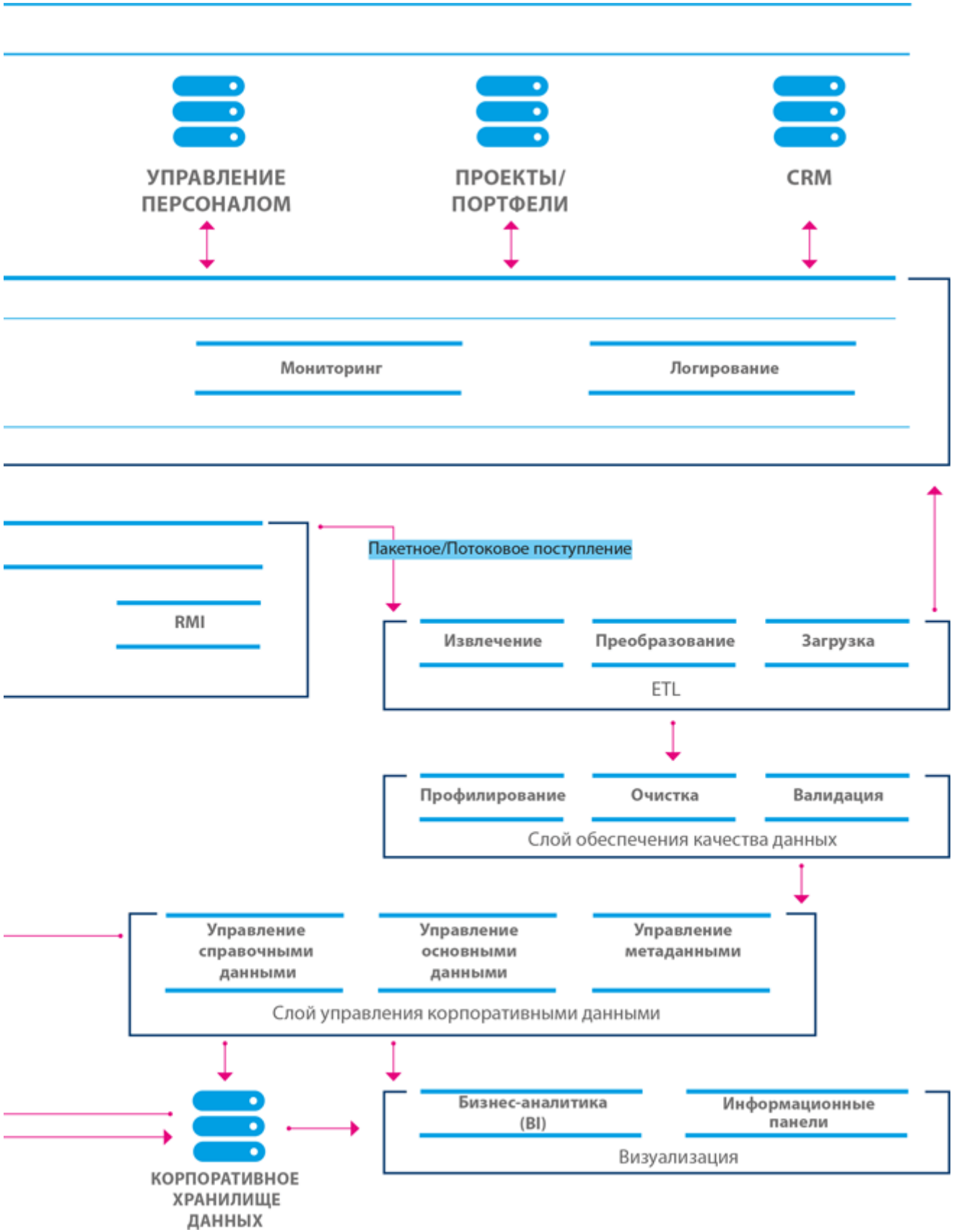


Рис. 12.3.

Представление целевой многоуровневой интеграционной архитектуры (пример)*



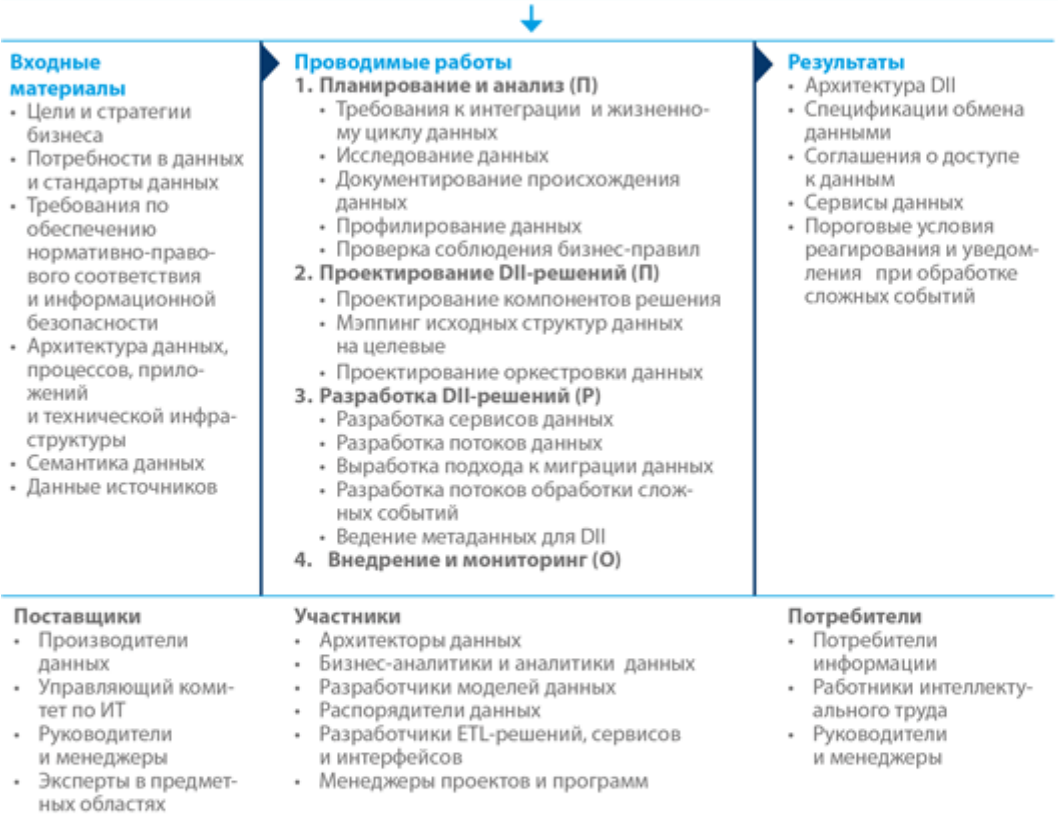
ИНТЕГРАЦИЯ И ИНТЕРОПЕРАБЕЛЬНОСТЬ ДАННЫХ

Определение: управление перемещением и консолидацией данных как внутри приложений и организаций, так и в рамках обеспечения их взаимодействия

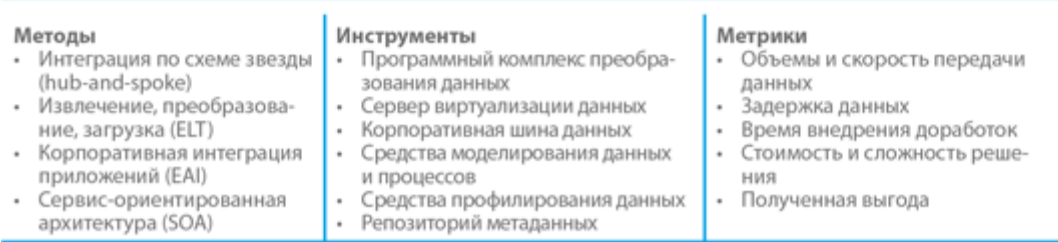
Цели:

- 1. Предоставление данных с соблюдением требований по обеспечению информационной безопасности и нормативно-правового соответствия в нужном формате и в заданные сроки
- 2. Снижение стоимости и сложности решений по управлению данными за счет разработки общих моделей и интерфейсов
- 3. Выявление значимых событий и автоматический запуск процедур выдачи уведомлений и принятия мер
- 4. Поддержка функций BI, аналитики, управления основными данными и обеспечение операционной эффективности

БИЗНЕС-ДРАЙВЕРЫ



ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 12.4.

Контекстная диаграмма. Интеграция и интероперабельность данных**

Совместное использование данных на базе общих определений, сервисов и протоколов лежит в основе осуществляемых организацией операций. Ключевые процессы и внутренние организационные структуры ориентированы на оптимальный обмен данными внутри и за пределами организации. Рациональное использование метаданных, справочных и основных данных гарантирует, что данные при обмене сохраняют свой смысл. Для упрощения обмена данными используются стандартные протоколы. Процессы автоматизированы.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Практики обеспечения интеграции и интероперабельности данных отслеживаются, измеряются и контролируются (на основе ключевых показателей эффективности) по всей организации. Общие определения данных и списки справочных кодов публикуются для обеспечения стандартизации и используются при совместном использовании данных. Передовые сервис-ориентированные подходы к обеспечению интероперабельности являются нормой.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Определены корпоративные политики и стандарты, устанавливающие стратегическую линию в отношении практики обеспечения интеграции и интероперабельности данных во всей организации. Взаимодействие между подразделениями осуществляется контролируемым образом. Определения данных, как правило, связаны с потребностями конкретных проектов. Используются признанные стандарты, такие как SOAP, REST API и т. п.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Для сокращения нерегламентированного обмена данными в организации введены формальные соглашения между подразделениями. Значительная часть обмена данными реализована по принципу «точка-точка». Совместное использование данных осуществляется, как правило, в соответствии с требованиями отдельных проектов. Общепринятые стандарты определения данных отсутствуют.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В тех сферах деятельности организации, где совместно используются общие данные, практически не применяются общие определения данных и общие сервисы (как внутри подразделений, так и при их взаимодействии и взаимодействии их систем). Значительная часть обмена данными осуществляется вручную, обычно по электронной почте в виде электронных таблиц и файлов данных (чему способствуют личные связи).

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 12.5.

Обобщенные характеристики уровней зрелости функции «Интеграция и интероперабельность данных»***

* Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

** DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

*** Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

12.2.5. Влияние на ценность данных

Внедрение современных интеграционных решений приносит дополнительные возможности косвенного влияния на повышение ценности всей совокупности корпоративных данных, в частности:

- Централизованная платформа интеграции обеспечивает ряд преимуществ, которые позволяют повысить операционную жизнеспособность организации за счет ограничения количества точек соприкосновения при передаче данных.
- Расширяются возможности по применению стандартов и более эффективному измерению степени соответствия требованиям руководства данными, поскольку может быть уменьшено количество точек мониторинга.
- Решение может упростить реализацию проектов за счет предоставления стандартных повторно используемых компонентов сервисов данных (которые в свою очередь могут предоставлять средства проверки качества данных).
- Диагностика и устранение проблем осуществляются более эффективно, поскольку в проекте интеграции задействовано меньше компонентов, обеспечивающих обмен данными между системами.
- Определять соглашения об уровне обслуживания и управлять их соблюдением гораздо легче, используя единую точку контроля и измерения.
- Информационный обмен между подразделениями и третьими сторонами упрощается, поскольку все данные проходят через централизованную сервисную платформу.
- Качество данных можно контролировать в процессе передачи напрямую, что облегчает выявление несоответствий в данных, используемых различными подразделениями организации.

12.3. Управление справочными и основными данными

В любой организации имеются совместно используемые данные, без которых трудно себе представить ее нормальную работу: юридические данные, клиентская база, сведения о поставщиках и контрагентах и пр. Пользователи этих данных рассчитывают на их согласованность в пределах организации, т. е. ожидают получить одну и ту же информацию об одном и том же объекте из разных источников. Разночтения и противоречия здесь порождают проблемы: задержки, коллизии, финансовые и имиджевые потери организации.

Подобного рода данные обычно подразделяются на справочные и основные.

Фундаментальную роль справочных и основных данных в информационном обеспечении процессов организации и важность поддержания высокого уровня их качества мы уже отмечали в главе 8.

12.3.1. Определение области знаний «Справочные и основные данные»

Справочные данные (reference data) – это любые данные, используемые для определения характеристик или классификации других данных, или же для соотнесения данных внутри организации с внешней информацией. В основном справочные данные состоят из кодов и их описаний (например, таблицы кодов и определений), но могут иметь и более сложную структуру, в том числе включать отображения и иерархии.

Типичным примером справочных данных являются общероссийские классификаторы технико-экономической и социальной информации. Согласно Федеральному закону от 29.06.2015 № 162-ФЗ «О стандартизации в Российской Федерации» это документы по стандартизации, распределяющие технико-экономическую и социальную информацию в соответствии с ее классификацией (классами, группами, видами и другим) и являющиеся обязательными для применения в государственных информационных системах и при межведомственном обмене информацией в порядке, установленном федеральными законами и иными нормативными правовыми актами Российской Федерации. К ним относятся общероссийские классификаторы валют (ОКВ), единиц измерения (ОКЕИ), стран мира (ОКСМ) и ряд других.

Для основных данных (master data) можно привести определение из ГОСТ Р ИСО 8000-2:2019:

«Основные данные: Данные, находящиеся во владении организации и описывающие объекты, которые являются независимыми и основными для этой организации и на которые нужно ссылаться при составлении транзакций.

Пример – Сообщение с кредитными картами относится к двум объектам, представленным в основных данных. Первый – это учетная кредитная карта счета в банке, идентифицированная номером кредитной карты. Основные данные, относящиеся к этой кредитной карте, включают в себя информацию по счету, требуемую банком-эмитентом. Второй – это коммерческий счет банка-получателя, идентифицированный номером, где основные данные включают в себя информацию об определенном торговце, требуемую банком-получателем.

Примечание 1. Как правило, основные данные описывают заказчиков, служащих, поставщиков, продукцию, пайщиков, услуги, инструменты, оборудование, а также правила и инструкции.

Примечание 2. Каждая организация определяет самостоятельно, какие данные следует считать основными».

Концептуально справочные и основные данные близки по своему назначению: и те и другие нужны для описания контекста транзакций, без которого невозможно создание и использование транзакционных данных (справочные данные при этом еще и определяют контекст для основных данных). Вместе они обеспечивают адекватное понимание данных.

Чтобы снизить издержки и риски, возникающие в результате рассогласования различных элементов справочных и основных данных, этими данными нужно управлять. Вопросы такого управления относятся к специальной области знаний – «Справочные и основные данные».

Важно иметь в виду, что и справочные, и основные данные – ресурсы совместного использования, управление которыми должно вестись исключительно на корпоративном уровне, а не на уровне отдельных систем⁴¹⁹.

12.3.2. Цели и бизнес-драйверы

Цели управления справочными и основными данными включают:

- обеспечение наличия в организации полных, согласованных, актуальных и достоверных основных и справочных данных по всему спектру процессов;
- обеспечение возможности совместного использования основных и справочных данных в рамках всех функций и приложений организации;
- снижение стоимости и сложности использования и интеграции данных за счет применения стандартов, общих моделей данных и шаблонов интеграции.

Самые распространенные драйверы управления справочными и основными данными⁴²⁰:

● **Выполнение требований организации к данным.** В различных областях работы организации требуются одни и те же наборы данных – и нужна уверенность в их полноте, актуальности и согласованности. Справочные и основные данные часто служат фундаментом при определении таких наборов данных (например, для планомерного и полного учета всех клиентов в аналитических выкладках необходимо четкое и последовательно применяемое определение клиента).

● **Управление качеством данных.** Противоречивые, некачественные или неполные данные приводят к неверным решениям и упущенным возможностям; управление справочными и основными данными позволяет снизить подобные риски за счет обеспечения полного и согласованного представления всех важных для организации сущностей.

● **Управление затратами на интеграцию данных.** Стоимость интеграции данных из новых источников в сложную информационную среду только повышается при отсутствии качественных справочных и основных данных, необходимых для минимизации разночтений в определениях критически важных сущностей.

● **Снижение риска.** Справочные и основные данные позволяют упрощать архитектуру обмена и

⁴¹⁹ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴²⁰ Там же.

совместного использования данных, снижая за счет этого издержки и риски, обусловленные избыточной сложностью ИТ-среды.

12.3.3. Специфика управления основными данными

Основные данные описывают ключевые бизнес-сущности (например, сотрудников, клиентов, продукты, финансовые структуры, ресурсы, адреса и т. д. и т. п.), определяющие контекст для бизнес-транзакций и их анализа. Сущность (entity) – это какой-либо объект реального мира (человек, организация, место или предмет). Сущности представлены своими экземплярами (entity instances), которые могут быть описаны в форме строк табличных данных или записей.

Основные данные требуют выявления и (или) выработки достоверной версии правды (trusted version of truth) для каждого экземпляра концептуальных сущностей, таких как продукт, место, счет, физическое лицо или организация, и поддержания этой версии в актуальном состоянии. Главная трудность при управлении основными данными связана с разрешением сущностей (entity resolution) – процессом определения различий и управления связями между данными различных систем и процессов. Экземпляры объектов, описываемых строками таблицы основных данных, в отдельных системах организации обычно представлены по-разному. В рамках управления основными данными должны быть отработаны механизмы разрешения этих рассогласованностей, иначе не получится однозначно и непротиворечиво идентифицировать одни и те же экземпляры каждой сущности (будь то клиенты, продукты и т. п.) в различных контекстах. Этим процессом необходимо управлять постоянно, чтобы не допустить рассогласования идентификаторов экземпляров сущностей основных данных на протяжении всего времени их использования.

Таким образом, управление основными данными (master data management, MDM) подразумевает контроль значений и идентификаторов, обеспечивающий их согласованность во всех системах и наиболее точное отражение актуальных сведений об основных бизнес-сущностях. Цели MDM включают обеспечение доступности точных текущих значений основных данных и минимизацию риска, связанного с их неоднозначной идентификацией (т. е. с появлением в системах идентификаторов, относящихся к нескольким экземплярам одной и той же сущности или соответствующих двум или более сущностям)⁴²¹.

12.3.4. Специфика управления справочными данными

Как уже отмечалось, справочные данные – это любые данные, которые используются для определения характеристик или классификации других данных, или же для соотнесения данных внутри организации с внешней информацией. Классификации могут, например определять статусы или типы (например, статус заказа: новый, обрабатывается, закрыт, отменен). Внешняя информация может включать данные о географическом местонахождении или применимых стандартах (и определяться, например, кодом страны).

Справочные данные отличаются от основных и транзакционных данных наличием только им присущих характеристик:

- справочные данные, как правило, менее изменчивы, чем другие виды данных, за некоторыми исключениями (например, данные об обменных курсах валют) они меняются нечасто;
- они обычно проще по структуре и менее объемны, чем наборы транзакционных или основных данных, т. е. таблицы справочных данных содержат меньше столбцов и меньше строк;
- никаких трудностей с разрешением сущностей при управлении справочными данными не возникает (в отличие от основных данных).

Управление справочными данными (reference data management, RDM) подразумевает контроль допустимых множеств значений данных и их определений. Цель RDM – обеспечить организации доступ к полному набору точных и актуальных текущих значений всех представляемых справочными данными понятий.

Одна из главных трудностей в управлении справочными данными – правильно определить их владельца, т. е. лицо, отвечающее за их определение и ведение. Часть справочных данных может поступать в организацию из внешних источников; другая часть – быть разбросанной по различным подразделениям и

421 Там же.

не иметь формального владельца; еще какие-то справочные данные могут генерироваться и учитываться в одном подразделении, а полученные значения использоваться в других подразделениях. Поэтому определение ответственных за сбор и обновление данных – важная функция RDM. Отсутствие распределения ответственности в сфере RDM порождает риск, поскольку разночтения в справочных данных влекут за собой неправильное понимание контекста данных (например, когда два бизнес-подразделения по-разному классифицируют одно и то же понятие)⁴²².

12.3.5. Контекстная диаграмма области знаний «Справочные и основные данные» и уровни зрелости соответствующих функций

Контекстная диаграмма области знаний «Справочные и основные данные» представлена на рисунке 12.6.

422 Там же.

СПРАВОЧНЫЕ И ОСНОВНЫЕ ДАННЫЕ

Определение: управление совместно используемыми данными, направленное на достижение целей организации, минимизацию рисков, обусловленных избыточностью данных, обеспечение повышения качества данных и снижение затрат на интеграцию данных.

Цели:

1. Поддержка совместного использования информационных активов в различных областях управления бизнесом и различными приложениями в масштабах организации
2. Предоставление доверенного источника согласованных справочных и основных данных проверенного качества
3. Снижение затрат на ведение и уменьшение сложности данных за счет использования стандартов, общих моделей данных и шаблонов интеграции

БИЗНЕС-ДРАЙВЕРЫ

Входные материалы	Проводимые работы	Результаты
• Бизнес-драйверы • Кросс-функциональные требования • Отраслевые стандарты • Глоссарий данных • Наборы данных и кодов из внешних источников (как коммерческих, так и открытых) • Бизнес-правила	1. Определение драйверов и требований (П) 2. Валидация определений данных (К) 3. Анализ и оценка источников данных (П) 4. Определение архитектурного подхода (Р) 5. Моделирование данных (Р) 6. Внедрение процессов распоряжения и ведения (К) 7. Определение руководящих политик (К) 8. Реализация совместного использования/сервисов интеграции (Р, О) • Определение источников данных • Публикация справочных и основных данных	• Требования к справочным и основным данным • Модели данных и шаблоны интеграции • Достоверные справочные и основные данные • Повторно используемые сервисы данных
Поставщики	Участники	Потребители
• Эксперты в предметных областях • Распорядители данных • Разработчики приложений • Поставщики данных • Бизнес-аналитики • Системные аналитики в области инфраструктуры	• Аналитики данных • Разработчики моделей данных • Распорядители данных • Специалисты по интеграции данных • Архитекторы данных • Аналитики качества данных	• Аналитики основных данных • Специалисты по интеграции данных • Архитекторы данных • Пользователи приложений • Разработчики приложений • Архитекторы решений

ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы	Инструменты	Метрики
• Соглашения об условиях использования • Ключевые перекрестные ссылки, необходимые для бизнеса • Ведение и анализ журналов обработки	• Средства моделирования данных • Репозитории метаданных • Инструменты профилирования и контроля качества данных • Средства интеграции данных • Платформы MDM-приложений • Архитектура совместного использования / Архитектура интеграции данных	• Показатели качества и соответствия данных • Интенсивность изменения данных • Интенсивность обращения к данным и сервисам данных • Доступность данных • Полнота реализации функций по распоряжению • Объемы совместно используемых данных

(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 12.6.

Контекстная диаграмма. Справочные и основные данные*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Проблема управления справочными данными связана с их использованием. Чтобы управление справочными данными было эффективным (обеспечивало актуальность и согласованность в рамках

различных приложений и областей применения), оно должно осуществляться с помощью технологии, позволяющей потребителям данных (сотрудникам и системам) оперативно получать к ним доступ в процессе совместной работы на многих платформах.

Как и в случае управления другими видами данных, управление справочными данными требует планирования и проектирования. Архитектура и модели справочных данных должны учитывать, как будет осуществляться хранение, ведение и совместное использование справочных данных. Поскольку это общий ресурс, он требует высокой степени развития практики распоряжения данными. Чтобы получить максимальную отдачу от централизованной системы управления справочными данными, организация должна разработать политику руководства, которая требовала бы использовать эту систему и не позволяла сотрудникам вести свои собственные копии наборов справочных данных. Это может потребовать определенных усилий в части управления организационными изменениями, поскольку заставить людей отказаться от своих электронных таблиц на благо организации не так просто.

Управление основными данными – еще более сложная задача. Она иллюстрирует фундаментальные проблемы работы с данными:

- во-первых, люди имеют различные представления об одних и тех же понятиях, и выработать консенсус бывает непросто;
- во-вторых, информация имеет свойство эволюционировать, и для систематического учета этих изменений требуются планирование, знание данных, а также технические навыки.

Любая организация, признавшая необходимость MDM, вероятно, уже успела столкнуться с массой сложностей, обусловленных наличием в ИТ-среде множества разнородных систем, которые получают вводные по различным каналам и сохраняют ссылки на сущности реального мира в различных форматах и местах. По причине естественного роста накапливаемых объемов разнородной информации, а также возможных слияний и поглощений, процессы, обеспечивающие MDM исходными данными, могут содержать различные определения одних и тех же сущностей, а также использовать различные критерии и стандарты качества данных. Из-за всех этих сложностей лучше подходить к внедрению единой системы MDM поэтапно, вводя ее поочередно в различных предметных областях. Начинать лучше с простой области с небольшим числом сущностей и атрибутов, а затем продолжать выстраивать систему MDM методом расширения.

Планирование управления основными данными включает несколько базовых этапов. В каждой предметной области нужно:

- выявить потенциальные источники, данные из которых обеспечат создание комплексного всестороннего представления сущностей основных данных;
- разработать правила, обеспечивающие точность сравнения и корректность слияния экземпляров сущности, оказавшихся идентичными;
- определить подход к выявлению некорректно распознанных как идентичные и необоснованно объединенных экземпляров, дополненный корректной процедурой восстановления исходных экземпляров сущности;
- определить подход к распространению прошедших тест на достоверность данных во все системы организации.

Реализация процесса управления, однако, не так проста. Нужно не только наладить управление данными в системе MDM, но и обеспечить их доступность для использования другими системами и процессами. Для этого требуется технология, позволяющая предоставлять данные и получать обратную связь. Она также должна быть подкреплена политиками, которые должны способствовать тому, чтобы системы и бизнес-процессы использовали общие значения основных данных, и не позволять им создавать свои собственные «версии истины».

На рисунке 12.7 представлены обобщенные характеристики уровней зрелости функции «Управление справочными данными».

На рисунке 12.8 представлены обобщенные характеристики уровней зрелости функции «Управление основными данными».

12.3.6. Влияние на ценность данных

Налаженное управление справочными и основными данными обеспечивает большое количество

преимуществ.

- Хорошо управляемые основные данные повышают организационную эффективность и снижают риски, связанные с различиями в структуре данных между системами и процессами.
- Создаются возможности для обогащения некоторых категорий данных. В частности, данные о заказчиках и клиентах могут быть дополнены информацией из внешних источников, таких как поставщики маркетинговых или демографических данных.
- Поскольку и справочные, и основные данные предоставляют контекст для транзакций, они оформляют и приводят в порядок транзакционные данные, вводимые подразделениями организации при выполнении операций (например, в системах CRM и ERP). Кроме того, они задают рамки анализа транзакционных данных.

В организации активно проводится планирование и тестирование вносимых в справочные данные изменений. Это позволяет всесторонне изучить их влияние на функционирование операционных и аналитических систем и распределить ресурсы в соответствии с предполагаемой ценностью изменений.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Управление изменениями справочных данных осуществляется на основе версионирования. Изменения и уведомления об изменениях идентифицируются и отслеживаются. Изменения группируются в отдельно развертываемые версии справочных данных и связанных с ними сопоставлений и преобразований.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Осуществляется централизованное управление справочными данными. Изменения в справочных данных операционных систем учитываются в центральных наборах справочных данных. Для обеспечения соответствия изменениям, происходящим в операционных системах, реализованы необходимые сопоставления и преобразования.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Управление справочными данными конкретных операционных систем осуществляется на плановой основе. При этом обеспечение сопоставлений и преобразований рассматривается как часть деятельности по изменению справочных данных, что позволяет избегать непредусмотренных воздействий изменений на функционирование конкретных аналитических системы.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации отсутствуют какие-либо свидетельства деятельности по управлению справочными данными на корпоративном уровне. Управление справочными данными, содержащимися в операционных системах, осуществляется исключительно для рабочего использования этими системами. Практически в любой аналитической системе имеется частное решение для сопоставления (мэппинга) справочных данных, облегчающее подстройку под их изменения.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 12.7.

Обобщенные характеристики уровней зрелости функции «Управление справочными данными»*

* Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

В организации осуществляется перенос унаследованных приложений с целью использования ими основных данных из единого доверенного источника (на основе применения единых общеорганизационных каналов доставки основных данных). Прилагаются усилия, направленные на устранение копий основных данных. Управление основными данными осуществляется на базе применения специализированного хранилища, а не с помощью бизнес-приложений.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

В организации реализована функция управления основными данными. Для мониторинга и оценки достижения целей используются метрики. Во всех новых приложениях и в управленческой отчетности используются основные данные из единого доверенного источника. Ведутся работы по переносу унаследованных приложений с целью предоставления им аналогичных возможностей.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

В организации действует централизованно управляемая программа MDM, поддерживаемая политиками и стандартами. Реализуется стратегия в отношении анализа имеющихся копий основных данных и их замены на данные из единого доверенного источника (для применения в существующих и новых приложениях, а также в управленческой отчетности). Например, имеются свидетельства комплексного моделирования основных данных и (или) использования инструментов MDM в рамках реализации стратегической программы в области бизнес-аналитики (BI).

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Имеются свидетельства того, что некоторые подразделения проводят работу по идентификации основных данных и предпринимают усилия по их анализу и сокращению количества копий. Например, есть свидетельства некоторой консолидации основных данных в рамках различных взаимодействующих систем или же осуществляется деятельность по оценке и приобретению инструментов MDM, но в основном для специально выделяемых проектов.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации отсутствуют какие-либо свидетельства наличия последовательной стратегии в отношении выявления, анализа, управления и сокращения количества имеющихся копий (реплик) основных данных, даже в рамках отдельных бизнес-функций или подразделений.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 12.8.

Обобщенные характеристики уровней зрелости функции «Управление основными данными»*

* Smith P.; Edge J.; Parry S.; Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.



ПРАКТИЧЕСКИЙ ПРИМЕР

Реализуя программу управления данными, компания «Телеком Дубль» решает ряд важных задач в рамках обеспечения доступности и обслуживания данных:

- продолжает развивать продукты в части применения облачных вычислений;
- переходит к использованию единой корпоративной сервисной шины;
- переходит к использованию единого платформенного решения по управлению основными данными.

В свою очередь проводимые мероприятия обеспечивают надежную базу для совершенствования всех направлений бизнеса.

Литература к главе 12

- Федеральный закон от 29.06.2015 № 162-ФЗ «О стандартизации в Российской Федерации». – URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102374687&intelsearch=%E7%E0%EA%EE%ED+%EE+%F1%F2%E0%ED%E4%E0%F0%F2%E8%E7%E0%F6%E8%E8+162-%F4%E7>.
- ГОСТ Р ИСО 8000-2:2019. Качество данных. Часть 2. Словарь.
- Кузнецов С. В., Кознов Д. В. Управление мастер-данными в рамках итеративного подхода // Онтология 2017170–184. – DOI: 10.18287/2223–9537–2021–11–2–170–184.
- Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.
- Berson A., Dubov L. Master Data Management and Data Governance, McGraw Hill; 2nd edition, 2010.

Глава 13. Управление основными данными: практика внедрения

В предыдущей главе нами была рассмотрена функциональная область «Справочные и основные данные». В этой главе, с учетом фундаментальной роли указанной области в построении цепочек ценности и поставок данных, мы более подробно остановимся на такой важной теме, как внедрение управления основными данными.

Обсуждая референтную модель управления цепями поставок (SCOR-модель), мы, в частности, говорили о том, что в ней есть специальный раздел, описывающий лучшие практики. Практика – это уникальный способ настройки процесса или совокупности процессов. Уникальность может быть связана с автоматизацией процесса, используемыми в процессе специальными навыками или технологией, уникальной последовательностью выполнения операций процесса или уникальным способом распределения процессов между организациями и их взаимодействия. Важность учета передовых практик подчеркивается и в «Сегодняшней повестке дня для совета директоров» (см. главу 7).

На сегодняшний день внедрение MDM не является типовой, проработанной задачей, поскольку организации, особенно крупные, имеют большое количество особенностей. В этой связи представляется крайне полезным изучение действующих в этой области практик.

В данной главе предлагается функциональная модель MDM, которая должна помочь в первичной оценке и согласовании работ в так называемых итеративных MDM-проектах на самых ранних стадиях. Внимание сосредоточено на создании ИТ-инфраструктуры по поддержке MDM, ее наладке и выполнении необходимых аналитических работ (очистка и консолидация данных, классификация и иерархизация и т. д.). В дальнейшем эти аналитические работы должны выполняться в организации на постоянной основе с помощью внедренной ИТ-инфраструктуры, и задачей-максимум является полная автоматизация этих работ. Ориентируясь на программно-аналитические аспекты MDM, мы здесь намеренно не касаемся вопросов изменения бизнес-процессов и практик работы с данными в организации, обучения персонала и пр.

В качестве иллюстрации описываемого подхода в главе представлено несколько реальных индустриальных MDM-проектов, выполненных с использованием предложенной модели⁴²³.

423 Кузнецов С. В., Кознов Д. В. Управление мастер-данными в рамках итеративного подхода // Онтология проектирования,

В завершающем разделе приведен пример архитектуры информационных систем крупной компании. Архитектура ориентирована на специфику телекоммуникационной отрасли и основана на комплексном MDM-решении, которое, в частности, может быть построено с применением предлагаемой в главе модели.

13.1. Две стратегии внедрения MDM

Известны две принципиально разные стратегии внедрения MDM: «сверху вниз» и итеративная стратегия⁴²⁴.

Стратегия «сверху вниз» подразумевает следующую последовательность действий:

- создание стратегической MDM-концепции для организации;
- формирование требований к MDM-решению;
- внедрение и доработка существующего на рынке MDM-инструментария;
- выполнение необходимых организационно-административных работ;
- эксплуатация и сопровождение MDM-решения.

Как правило, стратегия «сверху вниз» осуществляется посредством серии проектов, которые выполняются различными внешними компаниями.

Итеративная стратегия подразумевает внедрение MDM для решения конкретной задачи с дальнейшим наращиванием MDM-функционала и (или) реализацией MDM для других сегментов данных организации, т. е. для решения других задач. При этом внедрение MDM за пределами одного конкретного MDM-проекта зависит от многих условий, в том числе от оправдания ожиданий организации от данного проекта, от наличия других задач, свободных средств, заинтересованных людей и т. д. Именно такая цепочка MDM-проектов и дает название всей стратегии – итеративная стратегия внедрения MDM.

Первую стратегию можно сопоставить с технологией push: внедрение инновации происходит на основе некоторой передовой технологии, которая должна решить различные, в том числе и не известные на данный момент проблемы организации. Вторую стратегию можно сопоставить с технологией pull: драйвер внедрения инноваций – сама организация, точнее, определенные ее задачи. Не отвергая первой стратегии, авторы ориентируются на вторую, как менее рисковую и позволяющую достичь конкретных практических результатов в обозримые сроки.

При реализации MDM-проекта в рамках итеративной стратегии возникает задача перевода требований заказчика на язык MDM, получивший значительное развитие⁴²⁵. Если задача, имеющаяся у организации-заказчика, хорошо переводится в термины MDM, то для ее решения можно использовать имеющийся на рынке MDM-инструментарий⁴²⁶, что существенно сокращает затраты на такой проект. При этом на практике оказывается, что заказчик, как правило, не владеет MDM-терминологией и часто под видом MDM-проектов пытается представить проекты иного класса или заказать реализацию MDM-решения с нуля. Ошибки здесь приводят к коллизиям, растянутым срокам и денежным потерям.

13.2. Ключевые процессы MDM и архитектура MDM-решения

Управление основными данными ориентировано на сбор и накопление данных из различных информационных систем – источников данных, консолидацию данных и их распределение (доставку) информационным системам – потребителям данных.

Можно выделить следующие ключевые процессы MDM в организации:

- управление моделью данных;

2021. Т. 11, 2 (40): 170–184. – DOI: 10.18287/2223-9537-2021-11-2-170-184.

⁴²⁴ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴²⁵ Там же.

⁴²⁶ Walker S., Dayley A., Parker S., Hawker M. Magic Quadrant for Master Data Management // Gartner, 2021.

- сбор и накопление данных;
- проверка, стандартизация и обогащение данных;
- разрешение конфликтов данных (разрешение сущностей);
- использование данных⁴²⁷.

Практика MDM в организации должна быть поддержана специальным ИТ-решением, созданным и внедренным в организацию. Важнейшая составляющая MDM-решения – центральный репозиторий данных (хаб). В него собираются данные, являющиеся кандидатами в основные данные, они надлежащим образом обрабатываются и доставляются потребителям. Определены четыре варианта архитектуры хаба данных⁴²⁸.

1. **Индексная архитектура** : на хабе хранятся только соответствующие ссылки (индексы) данных; это актуально для данных, которые нельзя копировать или перемещать.

2. **Консолидирующая архитектура** : данные регулярно загружаются в хаб, обрабатываются, и при этом обеспечивается доступ потребителей к этим данным.

3. **Централизованная архитектура** во всем подобна предыдущей, но в этом случае дополнительно хаб выполняет задачу разового ввода данных, и далее все изменения данных делаются непосредственно на хабе (информационные системы попадают в разряд потребителей).

4. **Смешанная архитектура** реализует сочетание консолидирующей и централизованной архитектур для различных основных данных организации. Если какие-то фрагменты данных организации запрещено перемещать, то для них может использоваться индексная архитектура.

Существует большое количество готового программного инструментария по созданию MDM-решений. Прежде всего, это такие продукты как SAP MDG, Informatica MDM, IBM InfoSphere MDM, которые ориентированы на решение стандартных задач MDM. Однако разнообразие практических задач столь велико, что некоторые производители (например, «Юнидата» и пр.) предлагают программные «конструкторы», которые позволяют получить MDM-решения для конкретных потребностей организаций.

13.3. MDM-проекты

На практике проекты в области MDM часто принимают за обычные ИТ-проекты. И требуется определить, что задачи организации имеют ярко выраженную MDM-специфику, что позволяет применить готовый MDM-инструментарий, а также привлечь компании-интеграторы, специализирующиеся на таких проектах.

Запрос организации имеет MDM-специфику, когда заказчику необходим сбор, обогащение и консолидация данных из различных источников, а также выдача этих данных различным потребителям, использующим их далее для обеспечения бизнес-деятельности. Источники могут находиться как в самой организации, так и вне ее. Например, требуется обогатить данные о клиентах организации информацией, собранной в соцсетях. Требование наличия нескольких потребителей для созданных основных данных менее жесткое и иногда не выполняется. Как правило, в организации существует критический бизнес-процесс, для эффективного исполнения которого требуются качественные обогащенные и консолидированные данные из различных источников. Например, речь может идти о процедуре проверки новых клиентов или спорных транзакций в банке.

С другой стороны, организации может быть и не нужен MDM-проект. Во-первых, когда речь идет об обработке однородных данных – например, сформированных посредством ввода одним или несколькими операторами. Такие проекты могут потребовать: создания логической модели данных, валидацию и очистку данных, обеспечение доступа к данным в различных режимах и т. д. Но при этом отсутствует главная задача MDM – консолидация данных из разных источников. Если для такой задачи операторный ввод данных заменить автоматизированным поступлением тех же данных из разных источников, то она

⁴²⁷ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴²⁸ Walker S., Dayley A., Parker S., Hawker M. Magic Quadrant for Master Data Management // Gartner, 2021.

приобретает MDM-специфику. Во-вторых, «вне юрисдикции» MDM-проектов лежат запросы на реализацию сложного бизнес-функционала. Такой функционал должен быть вынесен из MDM-решения в отдельные информационные системы⁴²⁹, а MDM-проект заканчивается доставкой основных данных потребителям.

Следует отметить, что MDM-специфика должна быть основной в проекте. Если этой специфики нет или она лишь частично присутствует, но является не главной, то данный ИТ-проект не является MDM-проектом.

13.4. Состав MDM-решения

В результате реализации MDM-проекта у организации-заказчика появляется MDM-решение, которое включает в себя: MDM-систему (программную часть решения); новые регламенты по работе с данными; обученных сотрудников организации, которые умеют использовать MDM-систему в соответствии с новыми регламентами; налаженный и запущенный процесс управления основными данными. Последний пункт важен, поскольку все может иметься в наличии, но деятельность по MDM в организации отсутствует. Например, из-за нерешенных вопросов безопасности или противоречий внутренних регламентов, или в силу большой загруженности сотрудников, которые должны участвовать в MDM. Таким образом, MDM-решение – это осуществляющаяся в организации деятельность, налаженная и обеспеченная всем необходимым.

MDM-система представляет собой развернутое в организации готовое программное обеспечение, которое реализует основные функции MDM – хаб данных, консолидацию и пр. Главная часть этого ПО – базовый MDM-продукт⁴³⁰, а также дополнительный набор ПО для решения частных вопросов. Наличие многофункционального готового ПО, которое следует лишь настроить и развернуть у организации-заказчика, существенно снижает стоимость и риски MDM-проекта. Однако некоторую часть MDM-системы приходится дорабатывать в рамках проектных мероприятий, чтобы отразить особенности организации, которые не удается покрыть стандартным инструментарием⁴³¹.

13.5. Описание модели

После первой оценки потребностей заказчика необходимо провести их детальный анализ в терминах MDM⁴³². Для этого предлагается специальная функциональная модель. Она описывает типовое MDM-решение, включая в себя максимальный объем функциональности, с тем чтобы можно было выбрать необходимые компоненты, которые нужно реализовать в данной ситуации⁴³³⁴³⁴.

429 Там же.

430 Там же.

431 Здесь можно привести аналогию с рынком ERP-решений: вначале эти решения, как и MDM-решения сегодня, предназначались для больших организаций и также создавались на основе базовых ERP-продуктов со значительными доработками. Далее стали появляться специализированные решения для отдельных отраслей (энергетики, добывающих отраслей, логистики и пр.). Аналогично, в области MDM сейчас выделяются отдельные классы ПО, ориентированного на работу с конкретной предметной областью (например, Управление информацией о продуктах (Product Information Management, PIM)).

432 Там же.

433 Предложенную модель можно также назвать идеальной моделью То Ве – эта терминология принята в области структурного* и объектно-ориентированного анализа**, а также в области реинжиниринга бизнес-процессов***.

* Yourdon E., Constantine L. Structured Design: Fundamentals of a Discipline of Program and Systems Design. Yourdon Press, 1975.

** Jacobson I. Object-Oriented Software Engineering. ASM press, 1992.

*** Ould M. A. Business Processes: Modelling and Analysis for Re-Engineering and Improvement. Wiley, 1995.

434 Кузнецов С. В., Кознов Д. В. Управление мастер-данными в рамках итеративного подхода // Онтология проектирования, 2021. Т. 11, 2 (40): 170–184. – DOI: 10.18287/2223-9537-2021-11-2-170-184.

Для удобства использования модель представлена с помощью метафоры полного жизненного цикла основных данных и состоит из трех этапов: сбор, обработка и доставка данных. Эти фазы содержат функциональные компоненты, каждый из которых описывает блок работ по управлению основными данными. Таким образом, функциональные компоненты модели включают в себя работы по наладке MDM и работы, выполняемые в рамках дальнейшего функционирования MDM-решения.

Например, при реализации одной из главных компонент модели «Консолидация данных» нужно:

- выполнить работы по наладке: наладить ПО, поддерживающее соответствующее рабочее место аналитика, определить правила для разрешения конфликтов и выполнить первый раз консолидацию данных из источников данных организации;
- осуществлять консолидацию данных в рамках дальнейшего функционирования MDM, поскольку далее данные из источников будут продолжать поступать на хаб данных.

Предлагаемая модель ориентируется на внедрение в организации новой ИТ-системы на основе готовых инструментов, которые могут быть настроены и доработаны под особенности задач заказчика. Поэтому каждая компонента модели имеет программную и аналитическую части. Часть функционала компоненты выполняет соответствующее ПО, а часть – человек (аналитик). Работы по наладке компоненты целесообразно разделить на наладку/реализацию некоторого ПО и выполнение аналитических функций. Например, создание логической модели МД – аналитическая функция, а очистка данных – программно-аналитическая. В последнем случае речь идет о создании и программной реализации специальных правил очистки, которые применимы именно для этих данных, именно для этой организации, и применении этих правил, включая анализ результатов, возможно создание новых правил. При этом используется готовое ПО, но отдельные специальные правила, ориентированные на специфику данных организации, могут быть реализованы в виде дополнительного ПО, созданного в рамках MDM-проекта.

Основные пакеты модели

• **Сбор данных.** В этот пакет включены компоненты, отвечающие за идентификацию данных-кандидатов в основные данные, так называемых сырых данных, а также за их дальнейший анализ и предварительную обработку. Сюда же входит доступ к различным источникам данных.

• **Обработка данных.** В этот пакет включается функционал по созданию и хранению основных данных в хабе данных, включая создание и поддержку логической модели данных, а также выполнение классификации, иерархизации. В хаб поступают предварительно обработанные сырые данные, полученные из источников данных. Здесь они обрабатываются, становясь основными данными.

• **Доставка данных.** В этот пакет включены функциональные компоненты, отвечающие за доставку основных данных системам потребителей. Источники и потребители данных могут совпадать полностью или частично. При этом оказывается важным решение вопросов разделения прав доступа к данным, а также реализация различных режимов доставки данных потребителям. Выделяют следующие режимы доставки: пакетный, режим реального времени и подписочный режим.

Предложенная модель ориентирована на сценарий не только одноразовой загрузки данных, но и на их повторяющееся обновление в хабе с учетом поступления новых «сырых» данных из источников.

Пакеты и функциональные компоненты модели представлены на рисунке 13.1.

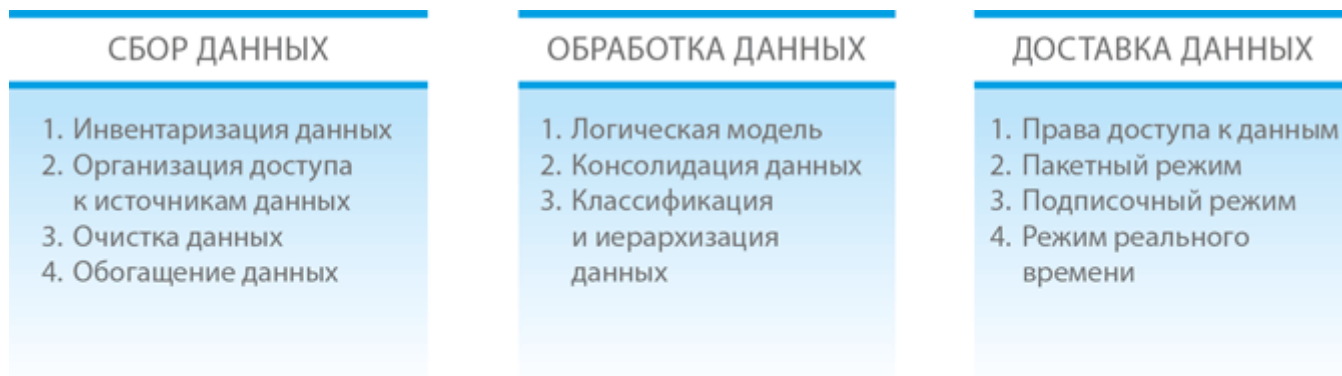


Рис. 13.1.

Схема функциональной модели типового MDM-решения*

* Кузнецов С. В., Кознов Д. В. Управление мастер-данными в рамках итеративного подхода // Онтология проектирования, 2021. Т. 11, 2 (40): 170–184. – DOI: 10.18287/2223–9537–2021–11–2–170–184.

13.5.1. Инвентаризация данных

В рамках этого пакета производится идентификация источников данных, а также определяется, какие именно данные из этих источников нужно преобразовывать в основные данные. Необходимо определить точный состав основных данных. Чем больше разнообразие данных, тем сложнее (и, следовательно, дороже) будет MDM-проект. При этом собирают лишь те атрибуты, которые будут востребованы потребителями данных. Важно провести типизацию данных, выяснить реальную заполняемость каждого существенного атрибута и его типовые значения. В этом пакете также определяется уровень доверия к различным источникам данных. Возможна ситуация, что некоторый источник имеет очень низкую степень доверия, в частности, он может хранить данные, которые давно не обновлялись. Обращаться к нему следует лишь в крайнем случае. Эта функциональная часть – преимущественно аналитическая.

13.5.2. Организация доступа к источникам данных

Поскольку сырые данные, которые должны использоваться для создания основных данных, находятся в различных источниках данных организации-заказчика, то для создания MDM-решения необходимо организовать программный доступ к этим данным. В большинстве случаев загрузка данных является многоэтапной процедурой и должна выполняться регулярно во время функционирования MDM-решения. Для автоматизации этой процедуры необходимо решить технические задачи, так как источники данных часто реализованы на разных платформах и могут не иметь программных интерфейсов доступа. Эта функциональная часть – преимущественно программная. Объем работ здесь во многом зависит от того, насколько обмен данными налажен в организации (например, уже могут быть внедрены технологии передачи данных между различными информационными системами организации).

13.5.3. Очистка данных

Под очисткой данных понимается устранение ошибок в данных и нормализация данных из различных источников перед их загрузкой в хаб. Это необходимо, поскольку в противном случае будет непросто искать в данных дубликаты, а также выполнять их консолидацию. Очистка данных – трудоемкий процесс. Первичная очистка, включая нормализацию и приведение значения всех атрибутов к единому формату, не является затратной, однако она значительно упрощает дальнейшие шаги по консолидации данных и восстановлению связей между сущностями.

Эта компонента – программно-аналитическая. На практике часто требуется программная доработка таких инструментов для корректной работы с различными форматами данных организации, либо реализация специальной логики по очистке информации. Например, источник данных может хранить несколько значений в одном атрибуте, и тогда требуется разделить эти значения по соответствующим полям. Для этого нужно разработать специальное ПО, которое выполнит это разделение.

13.5.4. Обогащение данных

Может оказаться, что для пользователей основных данных требуется дополнить имеющиеся в источнике данные, например, информацией из открытых источников.

Данная компонента – программно-аналитическая, поскольку требует анализа и разметки соответствующих данных из источника, а также реализации программного доступа к публичным источникам для обогащения данных.

13.5.5. Логическая модель

Эта компонента предназначена для создания и сопровождения логической модели основных данных. Такая модель должна отражать структуру консолидированных данных со всеми атрибутами, собранными из различных источников организации. Модель необходима для дальнейшей обработки основных данных, а также их доставки потребителям. Один из важнейших шагов при создании логической модели – восстановление/обнаружение различных связей в данных, которые отсутствовали в источнике, но появляются при консолидации.

Деятельность по созданию логической модели является аналитической. Она должна быть поддержана соответствующим ПО, включающим средства визуализации, перечисления атрибутов и связей между сущностями, а также программной связи созданной модели данных с соответствующим отражением ее элементов в источниках и (или) потребителях данных. При этом некоторые аспекты этого инструментария требуется дорабатывать под конкретный проект: например, в качестве источника и потребителя может выступать уже функционирующая информационная система, в которой модель данных жестко задана (типичный случай – ERP-система), и тогда доставка новых значений для существующих атрибутов будет требовать специальной программной реализации.

13.5.6. Консолидация данных

Эта функциональная компонента отвечает за загрузку данных из разных источников на хаб и выполнение консолидации реальных данных в соответствии с созданной логической моделью.

Процесс загрузки производится автоматически, с использованием соответствующих инструментов. При его выполнении возникают конфликты, которые разрешаются следующими способами.

- «Вручную» – эксперт предметной области разрешает конфликт; этот способ применяется для критических данных (например, юридических), где ошибки недопустимы и поэтому автоматические алгоритмы разрешения конфликтов неприемлемы.

- Семантический (онтологический) подход, который применяется для данных, которые хранятся в виде онтологий. Если фрагмент данных из источника попадает с другими фрагментами в одну онтологию, то эти фрагменты являются консолидированными.

- Методы искусственного интеллекта, в частности методы машинного обучения, которые обучаются на типичных ситуациях, чтобы разрешать возникающие в процессе консолидации конфликты автоматически.

- Смешанные стратегии – например, с помощью алгоритмов искусственного интеллекта экспертам представляется на одобрение предварительные варианты разрешения конфликтов. Такой подход может снизить трудоемкость процедуры разрешения конфликтов без снижения качества.

Загрузка данных из источника может осуществляться одноразово, например, в случае централизованной архитектуры хаба данных или при наличии источников, которые прекратили свою работу, но содержат ценные данные. Иначе помимо первичной загрузки требуется организовать регулярное обновление основных данных.

Данная компонента – программно-аналитическая. Программной частью является доработка ПО деятельности аналитика по консолидации данных для работы со специфическими данными, а также для реализации уникальных правил консолидации и правил разрешения конфликтов. Если используются алгоритмы искусственного интеллекта, то они должны быть адаптированы под конкретную задачу. Например, это могут быть обновляемые или самонастраиваемые правила для разбора конфликтов данных при консолидации.

13.5.7. Классификация и иерархизация данных

Организация нуждается в максимальной систематизации и упорядочении основных данных (например, основных активов, поставщиков, клиентов). Для этого требуется разделять их на группы и выделять значимые атрибуты для каждой группы. Необходимо связать данные организации с различными внешними классификаторами – государственными стандартами, отраслевыми классификаторами и т. д., а во многих случаях также иерархизировать данные. В качестве примера можно привести ситуацию, когда новый заказчик организации оказывается в том же организационном сегменте большой корпорации, что и предыдущий; в таком случае нет необходимости получать новое разрешение у службы безопасности, что экономит время и ресурсы. Деятельность по классификации и иерархизации данных производится над неструктурированными данными и может быть связана с обогащением данных.

Данная компонента – аналитическая в части создания правил иерархизации и классификации. Она должна быть поддержана соответствующим ПО, которое позволяет осуществлять отладку этих правил на малом количестве данных и далее применять их к полному объему данных. Для подсказок аналитику и создания полуавтоматической классификации данных все больше используют различные методы искусственного интеллекта, в частности методы машинного обучения.

13.5.8. Права доступа к данным

Потребители данных могут находиться в различных бизнес-контурах организации и иметь разные права доступа к данным. В рамках этой компоненты требуется определить стратегию прав доступа к основным данным и выполнить ее реализацию. Здесь необходимо полагаться на существующие в организации роли и связанные с ними права, взаимодействуя со службой безопасности организации.

Данная компонента включает работы, которые не являются трудоемкими и оказываются преимущественно аналитическими: политика разграничения прав доступа к основным данным реализуется средствами администрирования информационной системы. Однако создание соответствующей спецификации (какие данные и кому должны быть доступны) – ответственная работа, требующая глубокого знания данных и бизнес-процессов, а также структуры организации.

Таблица 13.1.

Характеристики MDM-проектов *

№	Аббревиатура	Название	Сектор индустрии
1	КМТР	Каталог материально-технических ресурсов	Энергетика
2	ПК	Продуктовый каталог	Телекоммуникации
3	КТУ	Каталог товаров и услуг	Транспорт
4	СКБ	Сегментирование клиентской базы	Розничная торговля
5	ЛКГ	Личный кабинет горожанина	Государственное управление
6	ПДК	Проверка данных клиента	Энергетика и тяжелая промышленность

* Кузнецов С. В., Кознов Д. В. Управление мастер-данными в рамках итеративного подхода // Онтология проектирования, 2021. Т. 11, 2 (40): 170–184. – DOI: 10.18287/2223–9537–2021–11–2–170–184.

13.5.9. Пакетный режим

Эта компонента отвечает за загрузку и обновление основных данных у потребителей в соответствии с некоторым расписанием. Многие потребители ориентированы на получение пакетных выгрузок данных в промежуточные базы («витрины данных»), с которыми они работают в своем режиме. При этом каждая

витрина использует свой фрагмент модели данных. Целесообразно реализовать отдельный механизм управления такими витринами для отслеживания своевременного обновления (получения ими актуальных данных), а также для журналирования запросов на получение данных разными потребителями. Таким образом отслеживается, какие именно данные используются теми или иными потребителями и в каком режиме; какие конфликты данных возникают в связи с теми или иными источниками и как это соотносится с потреблением данных.

Данная компонента имеет программную часть по наладке/реализации интерфейса MDM-системы с соответствующими потребителями. Аналитическая часть заключается в определении тех потребителей и тех частей основных данных, которые нуждаются именно в такой стратегии.

13.5.10. Подписочный режим

В рамках этого режима каждый потребитель подписывается на определенный фрагмент основных данных (часть логической модели или множество сущностей и их атрибутов). Далее формируется одна или несколько очередей, куда выгружаются наиболее актуальные данные после их очередного обновления. После этого все потребители считывают свои обновления из той или иной очереди согласно своей подписке. Сложность реализации подписочной модели состоит в том, что необходимо либо повторно использовать существующий механизм очередей, которым уже пользуются потребители, либо доработать соответствующие информационные системы для использования очередей MDM-решения.

Данная компонента – программно-аналитическая.

13.5.11. Режим реального времени

Данная компонента включает функционал по доставке основных данных потребителям в режиме реального времени, т. е. непосредственно после изменения данных. Такой режим часто трудно реализовать из-за конфликтов на стороне потребителя, поскольку соответствующая информационная система может временно блокировать доступ к фрагменту данных из-за выполнения некоторой операции, и это приведет к задержкам с обновлением записи в рамках MDM. Например, на сайтах телекоммуникационных компаний часто есть функция проверки возможности подключения той или иной услуги по адресу. Потенциальный клиент заходит на сайт компании, вводит свой адрес (возможно, с ошибками) и выбирает интересующую его услугу, например широкополосный доступ в интернет. Промедление с ответом сайта критично, т. к. клиент может уйти к конкуренту. Поэтому такая проверка адреса становится задачей поиска данного адреса (очистка от опечаток) и указанной услуги в режиме реального времени.

Данная компонента – программно-аналитическая.

13.6. Примеры MDM-проектов

Использование предложенной модели можно проиллюстрировать на примере MDM-проектов, выполненных при непосредственном участии авторов (см. табл. 13.1).

В приведенном далее описании выполненных проектов курсивом выделены компоненты функциональной модели, которые были в фокусе разработчиков.

КМТР. Проект направлен на создание системы для управления данными о материально-технических ресурсах крупной организации в энергетическом секторе. Система предназначалась для решения следующих задач: обеспечить качественными данными бизнес-процессы технического обслуживания, ремонта и управления запасами; консолидировать различную информацию за счет создания технологии стандартизации и унификации данных. В рамках проекта рассматривались данные о сырье и материалах, оборудовании, запасных частях и комплектующих изделиях, необходимых для обеспечения деятельности организации. К особенностям проекта можно отнести автоматизацию сложных регламентов организации по работе с информацией, затрагивающих более десяти различных подразделений, реализацию классификатора материальных ресурсов (*классификация и иерархизация данных*), построение логической модели данных.

ПК. Проект выполнялся для крупной телекоммуникационной организации и был нацелен на консолидацию информации по следующим направлениям: по продуктовым предложениям (услугам) для различных сегментов заказчиков; по проверке технических возможностей подключения услуг; по

объединению финансовой информации из систем биллинга и бухгалтерской отчетности. Основной акцент был сделан на инвентаризации данных о продуктах компании из различных источников, а также на создании единой логической модели основных данных с последующей консолидацией. Построено итоговое дерево продуктов компании с различными характеристиками, включая финансовые, для дальнейшего анализа отделом продаж и финансистами (*классификация и иерархизация данных*).

Таблица 13.2.

Описание индустриальных MDM-проектов в рамках предложенной функциональной модели

Элементы функциональной модели	Проекты					
	КМТР	ПК	КТУ	СКБ	ЛКГ	ПДК
Сбор данных						
Инвентаризация данных	Med	High	Med	Low	Med	High
Организация доступа к источникам данных	Low	Med	Low	Low	Low	Med
Очистка данных	Med	Low	Med	Med	Med	Med
Обогащение данных	N/A	Med	Med	High	N/A	Low
Обработка данных						
Логическая модель	High	High	Med	Low	Med	High
Консолидация данных	Med	High	High	High	Med	Med
Классификация и иерархизация данных	High	High	Med	N/A	N/A	High
Доставка данных						
Права доступа к данным	Low	Low	High	Low	High	Med
Подписочный режим	Med	Low	High	Low	High	Med
Режим реального времени	Med	Low	High	Low	High	Med
Пакетный режим	Med	N/A	Med	Med	High	High

КТУ. Основной задачей проекта была консолидация товаров и услуг, закупаемых крупной транспортной организацией. Было необходимо объединить информацию из различных классификаторов товаров и составить перечень услуг подрядчиков. Внутренним заказчиками этого MDM-решения стала служба закупки организации. В ходе проекта были идентифицированы по своему атрибутивному составу товары и услуги, имеющие различные цены. В результате были созданы монетарные метрики, т. е. подсчитана итоговая экономия организации по закупкам ввиду того, что требуемые товары и услуги стали закупаться по гарантированным минимальным доступным ценам автоматически. Фокус проекта был на *консолидации данных о закупаемых товарах, разграничении прав и обеспечении доступа к данным в рамках подписочного режима*.

СКБ. Проект разрабатывался для организации, занимающейся продажей модных товаров, и

предназначался для сегментирования клиентской базы и поддержки продаж в премиальном сегменте. Целью проекта было выявление клиентов из клиентской базы организации, которые активны в социальных сетях и имеют много подписчиков. Организация хотела заручиться их лояльностью с помощью дополнительных скидок и других мотивационных акций с целью получить больше потенциальных покупателей – подписчиков этих клиентов. В рамках проекта был сделан акцент на *обогащение и консолидацию данных*.

ЛКГ. Данный проект разрабатывался для городской государственной службы управления с целью создания умного личного кабинета горожанина. Требовалось выполнить интеграцию личного кабинета с многочисленными информационными системами федерального и регионального уровня с целью извлечения профильной информации о горожанине, например, сведений о его транспортных средствах, недвижимости или банковских счетах. Важными особенностями проекта была *информационная безопасность и разделение прав доступа к данным*, а также *получение основных данных в режиме реального времени и в рамках подписочной модели*.

ПДК. Проект разрабатывался для многопрофильной международной организации из сектора энергетики и тяжелой промышленности. Организация имеет сотни тысяч клиентов в разных странах мира, поэтому процедура формирования данных о новом клиенте оказывается трудоемкой. До создания MDM-решения она занимала 21 день, после – всего 8. MDM-решение позволило автоматизировать различные проверки, поиск конечных бенефициаров юридических лиц в корпоративных иерархиях, а также реализовать централизованный ввод информации. В рамках данного проекта основной акцент был сделан на *инвентаризации данных, создании логической модели, позволившей решить задачу поиска дубликатов юридических лиц и поиска аффилированных лиц*, а также *реализации доступа к основным данным в режиме реального времени с целью ускорить целевой бизнес-процесс*.

Для указанных проектов в таблице 13.2 показано, какие функциональные компоненты были реализованы в соответствующих MDM-проектах. При этом использовалась шкала:

- High – компонента является одной из основных в проекте, она бизнес-критична или технологически сложна;
- Med – компонента важна для проекта, но не является приоритетной или трудоемкой;
- Low – компонента реализована в облегченном варианте: она либо уже существует к началу проекта и требует лишь доработки, либо полная реализация компоненты вынесена в отдельный проект;
- N/A – данная компонента в рамках этого проекта не востребована.

13.7. Сопоставление существующих и описанного подходов

В главе 8 мы уже говорили о том, что за последние годы создано значительное количество референтных моделей и методологий управления данными в организациях^{435, 436, 437}.

Ряд методологий сфокусирован на ПО в сфере MDM⁴³⁸. Подход, описанный выше, отличается от этих методологий тем, что не зависит от конкретного базового MDM-продукта.

В работе О'Кейна и Морана⁴³⁹ предложена модель для построения системы MDM в организации, которая включает семь блоков: концепцию, стратегию, метрики, информационное управление, оргвопросы

⁴³⁵ CMMI Institute. Data Management Maturity (DMM)SM. Build, improve, and measure your Enterprise Data Management Program. ISACA, 2021. – URL: <https://cmmiinstitute.com/dmm>.

⁴³⁶ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴³⁷ IBM. IBM Data Governance Council Maturity Model. IBM Corporation, 2007. – URL: <https://ibm.co/2sRfBIIn>.

⁴³⁸ Informatica. Velocity Methodology. Best Practices. Informatica, 2008. – URL: <https://www.informatica.com/services-and-training/informatica-university/find-training/velocity-informatica-best-practices-methodology/ondemand.html>.

⁴³⁹ O'Kane B., Moran M. The Seven Building Blocks of MDM: A Framework for Success. Gartner. 2016.

и роли, ЖЦ информации, а также инфраструктуру. Эта модель предназначена для ранних стадий внедрения MDM, однако она ориентирована на стратегию «сверху вниз», покрывая всю деятельность организации по внедрению MDM. Функциональная модель, предложенная в данной главе, предназначена для использования при реализации итеративной стратегии, ориентированной на удовлетворение конкретных потребностей организации, которые выражаются в терминах MDM.

В свою очередь, в труде Мартина Офнера и его коллег⁴⁴⁰ предлагается модель для анализа жизненного цикла основных данных в организации с целью определить недостающие виды деятельности. Основными компонентами модели являются: портфолио данных; проектирование данных и системы; управление данными; поддержка данных. Эта модель слабо связана с программной частью MDM-решения, а также не рассматривает уникальные задачи организации по внедрению MDM.

В качестве дальнейшего развития предлагаемого подхода планируется детальная разработка методик оценки функционала MDM-проектов на ранних стадиях, а также создание детальных метрик сложности MDM-решений.

Кроме того, предполагается выполнить перевод (отображение) функциональности типового MDM-решения на различные MDM-продукты, а также осуществить более тесную интеграцию подхода с областью управления знаниями^{441, 442}.

13.8. Пример архитектуры информационных систем, основанной на комплексном MDM-решении

Одной из самых требовательных к инструментарию работы с данными групп заказчиков являются телекоммуникационные компании, которым необходимо на постоянной основе обеспечивать своих абонентов целым перечнем телекоммуникационных и ИТ-услуг и связывать огромный поток данных в рамках единого стройного механизма взаимодействия инфраструктуры и программных продуктов.

Даже обобщенная высокоуровневая схема поражает воображение грандиозностью и сложностью как самих систем, так и их взаимодействия друг с другом. Еще более сложен процесс общения этих систем на уровне данных, объем которых даже оценить непросто. В отдельных архитектурных сегментах создаются, накапливаются и обновляются столь разнородные и специализированные сведения, что зачастую только специалисты в каждой конкретной системе смогут ответить, с какими данными и по каким алгоритмам эти системы живут.

Попробуем разобраться, как же все-таки устроена жизнь данных в компаниях, принадлежащих к одной из самых сложных и высокотехнологичных отраслей постиндустриальной экономики, являющейся пионером в предоставлении клиентам сервисной модели.

Принципиальное значение имеет прежде всего выделение информационных источников, от управления которыми зависит весь процесс управления данными, как основными, так и транзакционными. Необходимо отметить, что, учитывая объем обрабатываемых сведений и важность привязки транзакционных данных к основным, для обеспечения большинства процессов телекоммуникационных компаний целесообразно рассматривать внедрение единого платформенного решения по управлению данными в масштабах всей организации, поскольку данные из различных систем (в первую очередь основные данные) сильно переплетены и зависимы друг от друга. Обеспечив работу с данными в одном месте (в среде единого инструментария) возможно обеспечить эффективное управление ими и, следовательно, большинством бизнес-процессов различных подразделений.

Рассмотрим домен (предметную область) «Клиентский каталог» (рис. 13.2). Соответствующая ему

440 Ofner M., Strauß K., Otto B., Osterle H. Management of the master data lifecycle: a framework for analysis // J. Enterp. Inf. Manag., 2013. 26 (4): 472–491.

441 Гаврилова Т. А. Представление знаний в экспертной диагностической системе АВТАНТЕСТ // Известия Академии наук СССР. Техническая кибернетика. 1984. 5: 165–173.

442 Гаврилова Т. А., Кудрявцев Д. В. Управление знаниями: от слов к делу / Т. А. Гаврилова, Д. В. Кудрявцев // Intelligent Enterprise: RE (Корпоративные системы). 2004. 12–13 (101): 48.

подсистема, функционирующая на базе единой платформы, обеспечивает управление данными клиентов, поддерживая их корректность, интероперабельность, чистоту, полноту, формат и другие требования к данным для работы с ними в информационных системах. Подсистема интегрирована с информационными системами, в которых имеются данные о клиентах в различных «разрезах» и представлениях, что позволяет платформе управления данными собрать все эти данные, провести их обработку, обогатить основную запись дополнительными сведениями (адрес, модели устройств, тип предоставляемых услуг и т. д.) и «опубликовать» – сохранить «золотую запись» в платформе и обеспечить ее распространение в информационные системы-получатели (они же могут быть и источниками, в таком случае данные будут актуализированы).

Рис. 13.2.

Подсистема «Клиентский каталог»

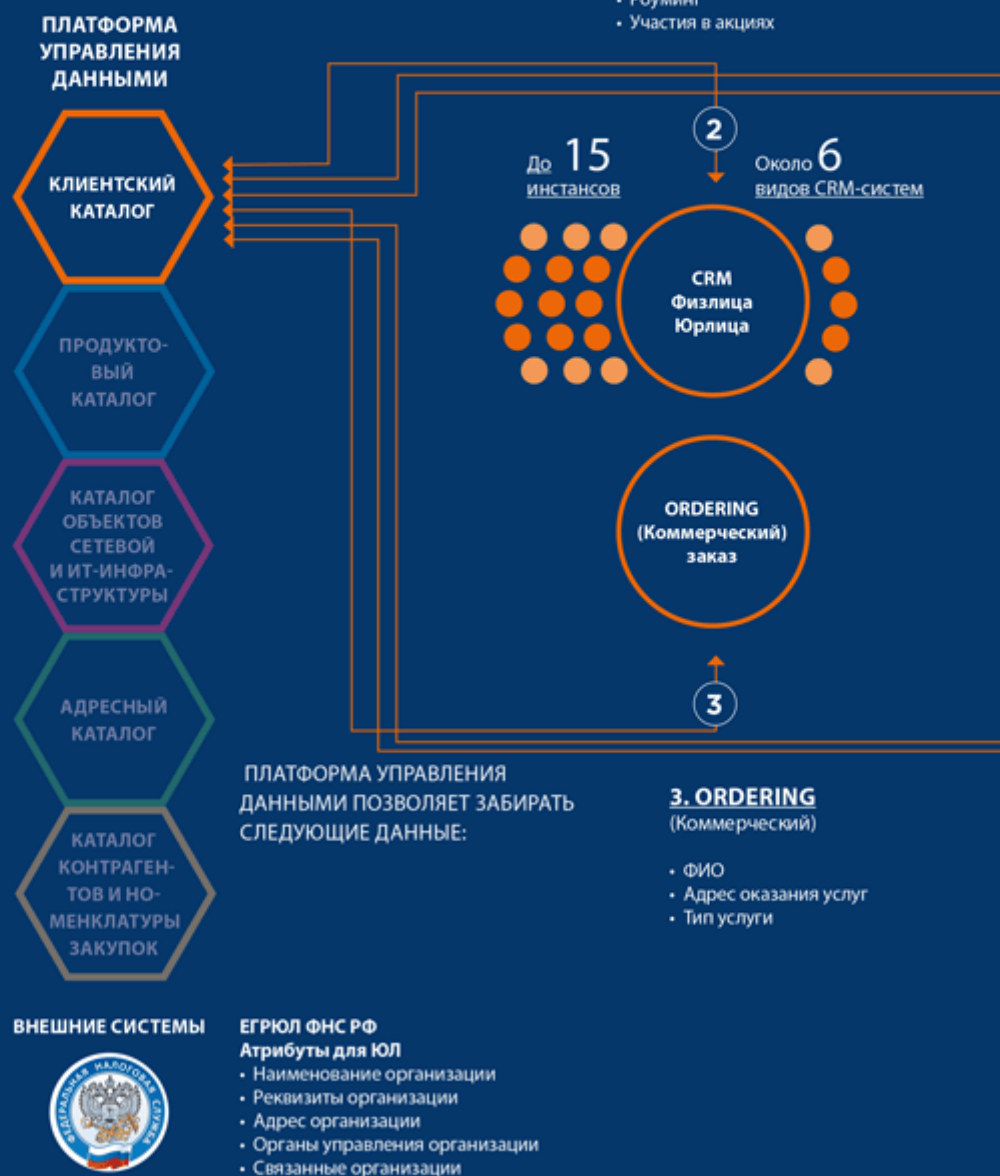
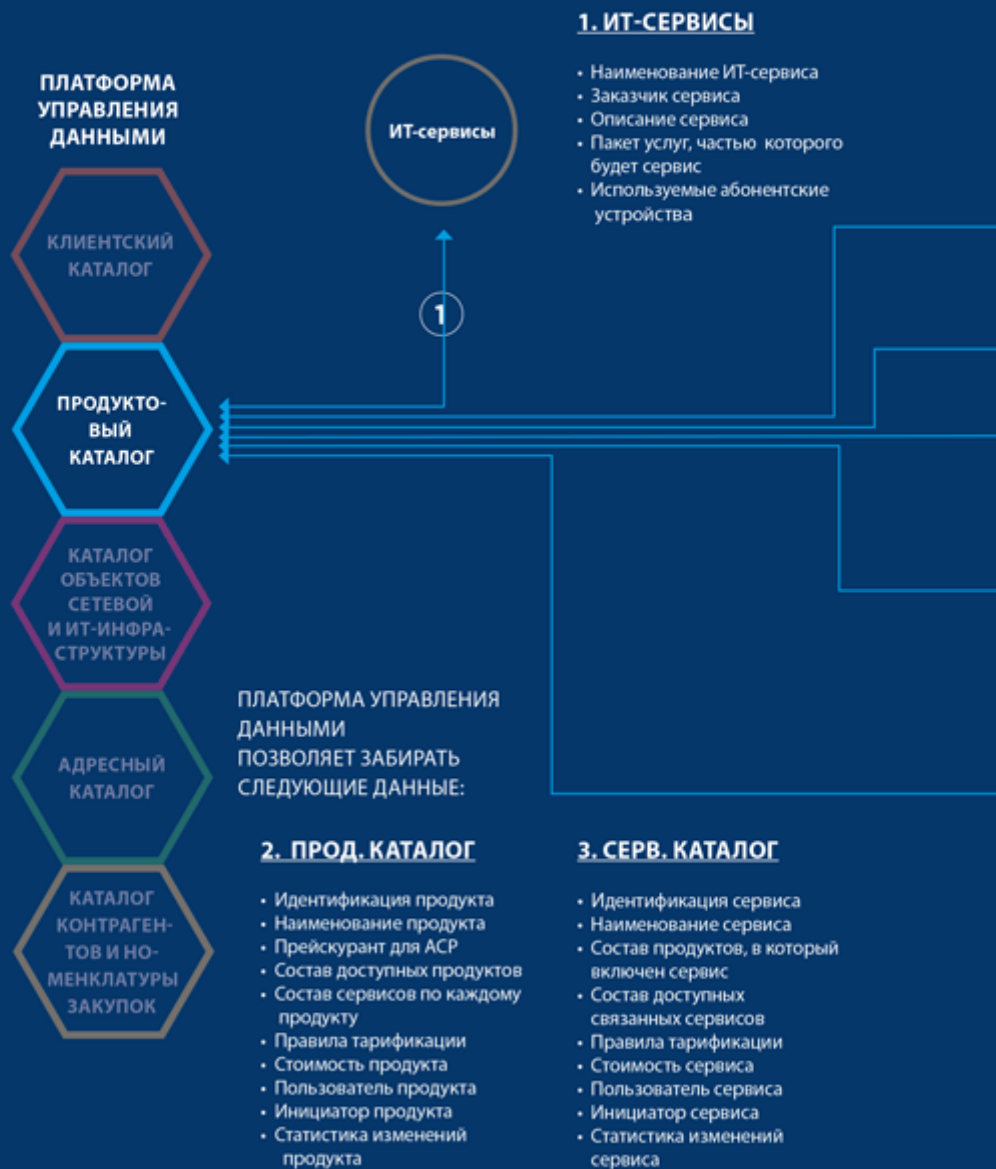




Рис. 13.3.

Подсистема «Продуктовый каталог»



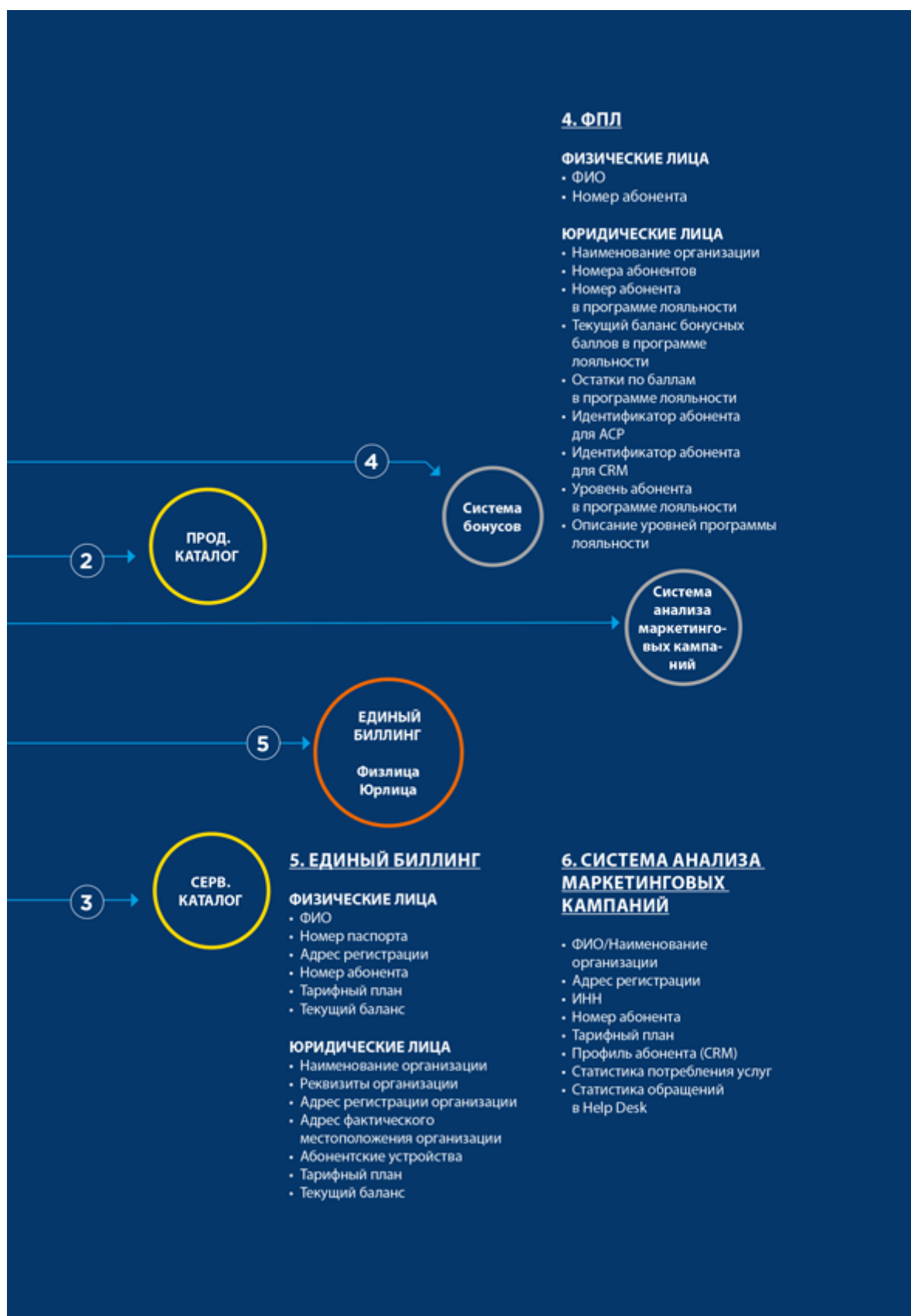


Рис. 13.4.

Подсистема «Каталог объектов сетевой и ИТ-инфраструктуры»



2. NRI

- Перечень поддерживаемого оборудования (производитель, поставщик, модель, версия ПО, дата поставки, дата включения, дата окончания гарантийной поддержки от производителя)
- Статус доступности
- Сетевая иерархия
- Статистика загрузки оборудования
- Статистика отказов/ремонтов оборудования

3. ORDERING

Технический

- Наименование оборудования
- Модель оборудования
- Номер оборудования
- Задействованные порты
- Настройка портов
- Доступные порты
- Связь с АСР

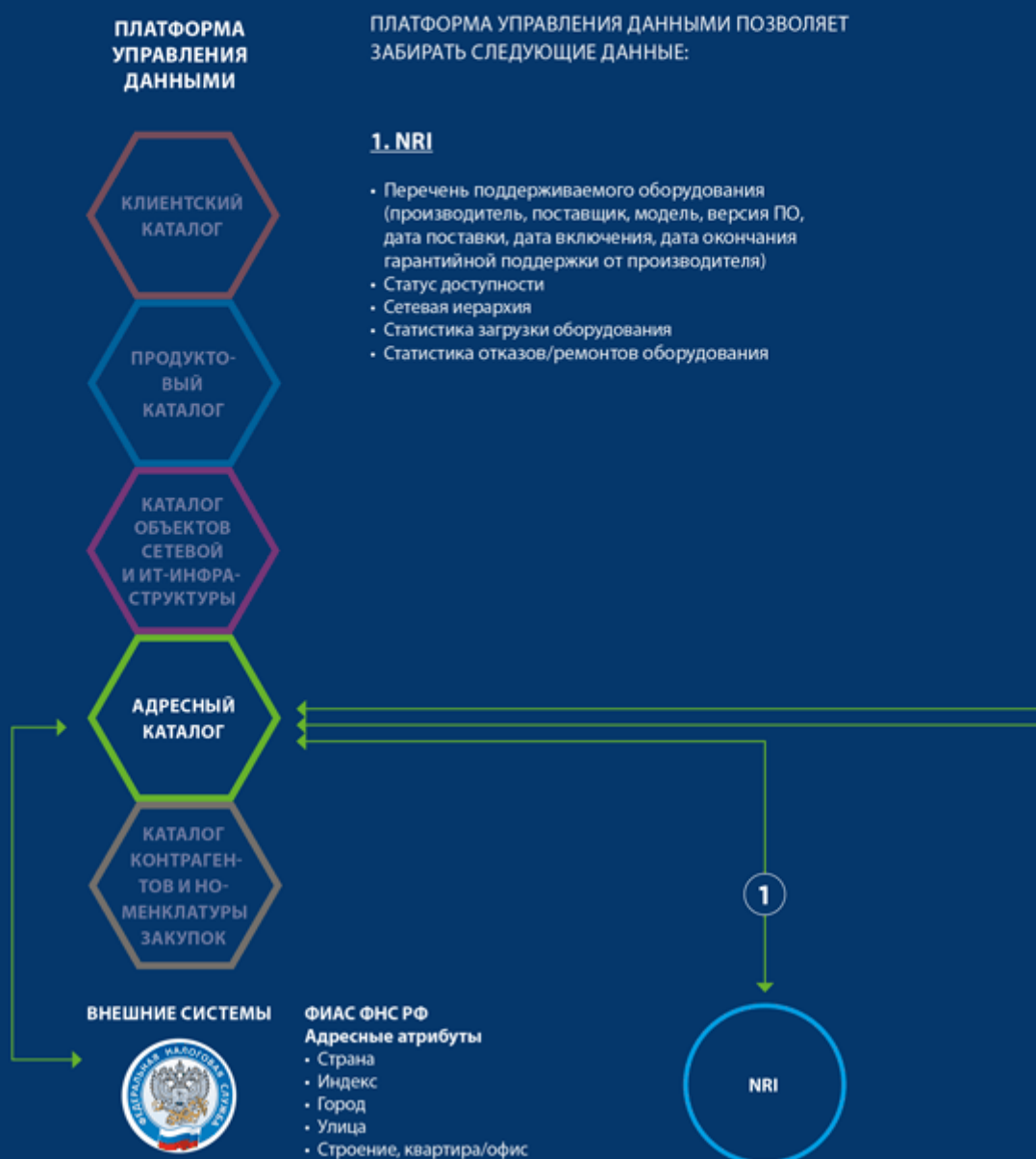
4. СКЛАДСКОЙ УЧЕТ

- Наименование склада
- Адрес склада
- Транспортная доступность склада (ж/д транспорт, авиатранспорт, автомобильный транспорт, курьерская доставка)
- Производитель оборудования
- Поставщик оборудования
- Модель
- Версия ПО
- Серийный номер оборудования
- Дата поставки
- Номер договора поставки
- Текущий статус оборудования (на складе/на сети/ в аренде/в ремонте/ возвращено производителю по рекламации)
- Дата включения
- Дата окончания гарантийной поддержки от производителя
- Статистика отказов



Рис. 13.5.

Подсистема «Адресный каталог»



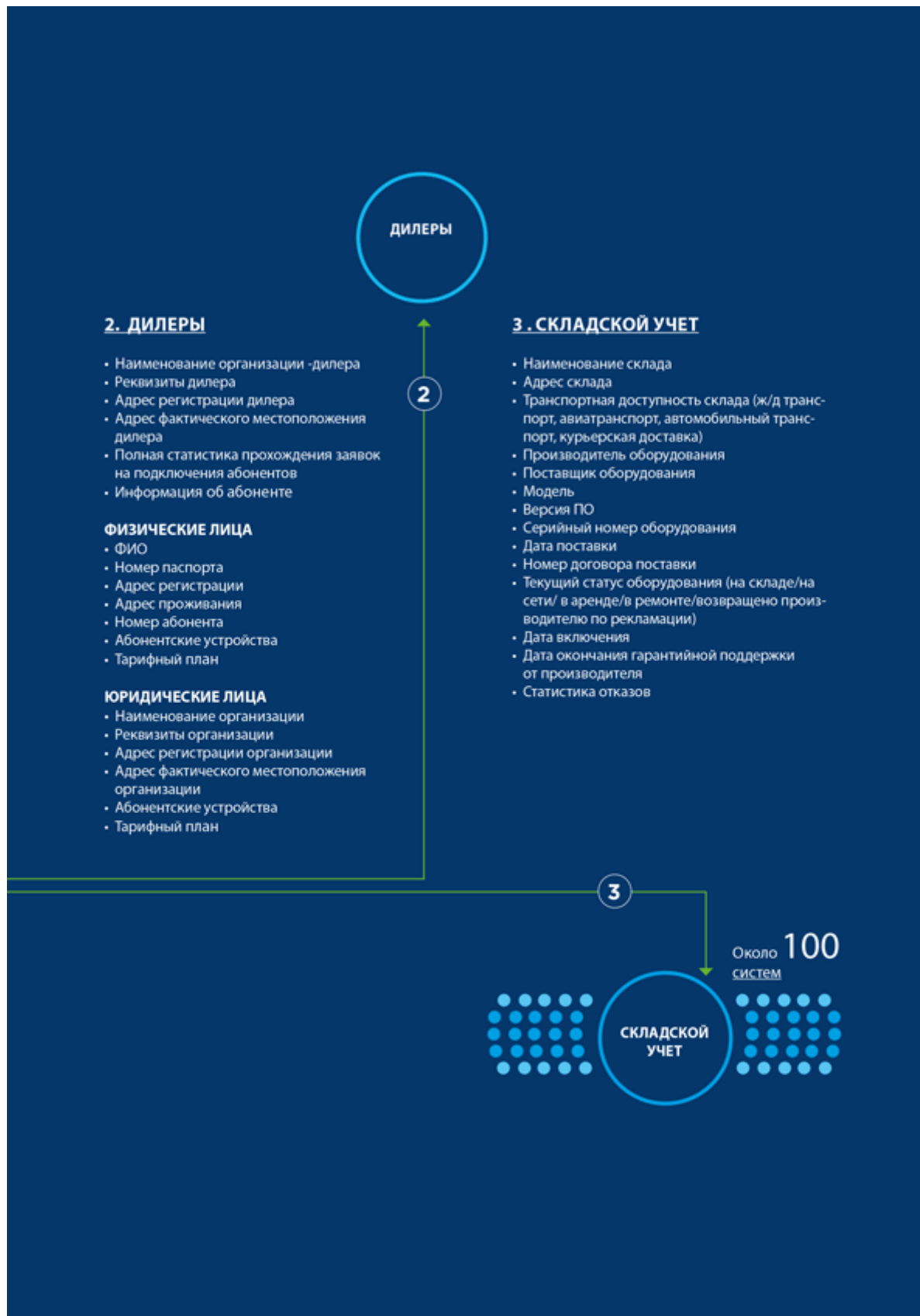


Рис. 13.6.

Подсистема «Каталог контрагентов и номенклатуры закупок»



2. NRI

- Перечень поддерживаемого оборудования (производитель, поставщик, модель, версия ПО, дата поставки, дата включения, дата окончания гарантийной поддержки от производителя)
- Статус доступности
- Сетевая иерархия
- Статистика загрузки оборудования
- Статистика отказов/ремонтов оборудования

3. СКЛАДСКОЙ УЧЕТ

- Наименование склада
- Адрес склада
- Транспортная доступность склада (ж/д транспорт, авиатранспорт, автомобильный транспорт, курьерская доставка)
- Производитель оборудования
- Поставщик оборудования
- Модель
- Версия ПО
- Серийный номер оборудования
- Дата поставки
- Номер договора поставки
- Текущий статус оборудования (на складе/на сети/ в аренде/в ремонте/возвращено производителю по рекламации)
- Дата включения
- Дата окончания гарантийной поддержки
- от производителя
- Статистика отказов

4. OEBS

- Контрагенты
- Номера договоров
- Условия договоров
- Стоимость оборудования
- и услуг
- Гарантийные обязательства
- Наименование участков сети (последние мили/ каналы связи)
- Постгарантийная поддержка
- Производители
- и поставщики оборудования
- Модели оборудования
- Серийные номера оборудования
- Версии программного обеспечения



Остановимся коротко еще на нескольких подсистемах на базе платформы управления данными, отвечающих за основные домены телекоммуникационной компании.

Подсистема «Продуктовый каталог» реализует функциональность работы с основными данными, на основе которых предоставляются услуги, обеспечиваются расчеты с абонентами, а также предоставляется возможность оперативного изменения тарифных планов (рис. 13.3).

Подсистема «Каталог объектов сетевой и ИТ-инфраструктуры» обеспечивает сбор и работу с данными, позволяющими управлять всеми техническими объектами, посредством которых предоставляются телекоммуникационные сервисы. Чаще всего это сложное технологическое оборудование, состоящее из различных технологических подсистем. Подсистемы, работающие с данными этого домена, обеспечивают полное описание оборудования в сети и на складах оператора, поддержку данного оборудования, описание всех неисправностей и ремонтов, произведенных силами оператора или подрядчиков, информацию о сроках и наименовании гарантийной поддержки и многое другое (рис. 13.4).

«Адресный каталог» – подсистема, поддерживающая качество данных об адресах абонентов, дилеров, нахождения оборудования, складов, а также детальное описание оборудования и предоставляемых услуг по каждому конкретному адресу (рис. 13.5).

«Каталог контрагентов и номенклатуры закупок» – отдельная интерпретация решения «Единого каталога товаров, работ и услуг» с гораздо более широким функционалом, позволяющим не только обеспечивать данными подсистемы закупок, но и работать с данными из информационных систем поддержки оборудования (учитывающих множество параметров устройств, как работающих «на сети», так и находящихся на складе). Каталог также формирует «единое пространство данных» для производственных систем и систем экономического учета и бюджетирования (рис. 13.6).

Отдельно следует упомянуть, что платформы управления данными могут быть реализованы как для работы исключительно операторов данных, так и иметь свои интерфейсы для работы с отдельными категориями пользователей, предоставляющие возможность оперативного отображения максимально широкого состава данных по конкретной предметной области. В качестве примера на рисунках 13.7 и 13.8 представлены подобные интерфейсы клиентского каталога телекоммуникационного оператора.

ИНТЕРФЕЙСЫ ПЛАТФОРМЫ УПРАВЛЕНИЯ ДАННЫМИ



Интерфейс оператора данных



Интерфейс клиентского каталога

**Рис. 13.7.**

Пример интерфейсов платформы управления данными

ИНТЕРФЕЙСЫ ПЛАТФОРМЫ УПРАВЛЕНИЯ ДАННЫМИ



Интерфейс оператора данных



Интерфейс клиентского каталога

**Рис. 13.8.**

Пример интерфейсов платформы управления данными



ПРАКТИЧЕСКИЙ ПРИМЕР

Опыт реализации MDM-проектов в «Телеком Дубль» показал целесообразность использования итеративного подхода, описанного в этой главе.

Не вызывает сомнения и оптимальность применения архитектуры, основанной на

комплексном MDM-решении. В частности, если говорить о клиентском домене, система управления основными данными содержит все сведения о клиенте в виде «золотых записей». Сюда относится информация как о самом клиенте, так и об использованных им тарифах и услугах, всех договорах, нормативных актах, сервисах и командах, которые этими сервисами управляют. Имея под рукой сведения такого объема и уровня детализации, «Телеком Дубль» может, например, формировать максимально приближенные к реальности KPI для каждого сотрудника.

Литература к главе 13

- Андриченко А. Н. Тенденции и состояние управления справочными данными в машиностроении // Онтология проектирования, 2012. 2 (4): 25–35.
- Голубев С. С., Лоцманов А. Н., Кузин А. Ю., Соловьев В. Г., Козлов А. Д., Григорьев Б. А. Отраслевая система государственной службы стандартных справочных данных нефтегазового комплекса // Законодательная и прикладная метрология. 2020. 3: 12–16.
- Немцов Э. Ф. ИСУЖТ и нормативно-справочные данные // Автоматика, связь, информатика, 2020. 2: 15–18.
- Чигиринский Ю. Л. Методика повышения надежности справочных данных // Известия Волгоградского государственного технического университета, 2011. 13 (86): 55–61.
- Янченко Г. А. К вопросу о стандартизации справочных данных плотностных свойств горных пород // Горный информационно-аналитический бюллетень (научно-технический журнал), 2011. 8: 111–115.
- Khatri V., Brown C. Designing data governance // Communications of the ACM, 2010. 53 (1).
- Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.
- Silvola R., Jaaskelainen O., Kropsu-Vehkapera H., Haapasalo H. Managing one master data – Challenges and preconditions // Industrial Management & Data Systems, 2011. 111 (1): 146–162.
- Zmud R. W. An Examination of ‘Push-Pull’ Theory Applied to Process Innovation in Knowledge Work // Management Science, 1984. 30 (6): 727–738.

Глава 14. Обеспечение доступности и обслуживание данных: развитие

В главе 12 мы начали обсуждение группы областей знаний по управлению данными (или функций управления данными), которые отвечают за укрупненную фазу их жизненного цикла, связанную с обеспечением доступности и обслуживания (рис. 9.4). Мы уже рассмотрели первые три функции, закладывающие основу способностей организации по управлению данными на этой фазе. В этой главе мы обсудим оставшиеся функции, которые можно рассматривать как обеспечивающие дальнейшее развитие этих способностей:

- ведение хранилищ данных;
- управление документами и контентом;
- хранение больших данных.

В отношении функций «Ведение хранилищ данных» и «Хранение больших данных» нужно заметить следующее.

DMBOK2 рассматривает ведение хранилищ данных в рамках более широкой области знаний «Ведение хранилищ данных и бизнес-аналитика». Поэтому далее, чтобы не было разночтений со сводом знаний по управлению данными, мы будем придерживаться такого же подхода. В этой главе мы начнем обсуждать область «Ведение хранилищ данных и бизнес-аналитика», уделяя основное внимание вопросам ведения хранилищ данных. Далее, в главе 17, посвященной использованию данных, обсуждение будет продолжено в части бизнес-аналитики.

Аналогичным образом мы будем рассматривать и функцию «Хранение больших данных», которая в

DMBOK2 разбирается в рамках общей темы «Большие данные и наука о данных»⁴⁴³. В этой главе мы начнем обсуждение данной темы, уделяя основное внимание вопросам хранения больших данных, а в главе 17 более подробно поговорим о науке о данных.

14.1. Ведение хранилищ данных и бизнес-аналитика

Понятие хранилище данных (Data Warehouse, DW) появилось в 1980-х годах для обозначения технологии, позволяющей организациям интегрировать данные из множества разнородных источников в рамках единой модели. С тех пор, особенно в связи с одновременным развитием бизнес-аналитики (Business Intelligence, BI) как основного драйвера принятия бизнес-решений, корпоративные хранилища данных успели стать обыденной вещью.

В главе 7, сопоставляя элементы референтной модели управления цепями поставок (SCOR-модель) с цепочкой поставок данных, мы отметили, что ведение хранилищ данных можно включить в группу процессов доставки, в частности складирования. При этом бизнес-аналитика больше соотносится с группой процессов «Делать», которая обеспечивает превращение материалов (данных) в различного рода информационные продукты.

14.1.1. Определение области знаний «Ведение хранилищ данных и бизнес-аналитика»

Хранилище данных (DW) включает два ключевых компонента – интегрированную базу данных, необходимых для принятия решений, и увязанное с ней программное обеспечение, используемое для сбора, очистки, преобразования и хранения данных из разнообразных внутренних и внешних источников. Кроме того, для поддержки функций ведения учета исторических данных, операционного и бизнес-анализа хранилище данных может включать вторичные витрины данных, т. е. выборочные копии данных из основного хранилища. В самом широком контексте под хранилищем данных может пониматься весь комплекс хранилищ, баз и витрин данных, используемых в организации в целях бизнес-аналитики.

Корпоративным хранилищем данных (Enterprise Data Warehouse, EDW) называют централизованное DW, предназначенное для информационного обеспечения BI-потребностей всей организации. EDW поддерживает корпоративную модель данных, что обеспечивает согласованность данных, используемых для принятия решений в масштабах организации.

Ведение хранилища данных включает осуществление текущих операций по извлечению, очистке, преобразованию, контролю и загрузке, обеспечивающих поддержку данных в хранилище в надлежащем состоянии. В процессе ведения DW первоочередное внимание уделяется обеспечению целостности и преемственности данных в историческом и бизнес-контекстах за счет применения к операционным данным адекватных бизнес-правил и реляционных связей. Кроме того, к сфере ведения DW относится также и поддержка процессов взаимодействия и согласования DW с репозиториями метаданных.

Понятие бизнес-аналитики (BI) имеет два смысловых значения. Во-первых, это вид анализа данных, который нацелен на изучение деятельности организации и выявление возможностей для развития бизнеса. Результаты такого анализа используются для совершенствования работы организации и достижения успехов в бизнесе. Во-вторых, под бизнес-аналитикой понимается еще и комплекс технологий, используемых для такого анализа данных⁴⁴⁴. Являясь логическим развитием инструментов поддержки принятия решений, инструменты бизнес-аналитики предоставляют возможности по формированию и обработке запросов (querying), извлечению информации (data mining), проведению статистического анализа (statistical analysis), формированию отчетности (reporting), сценарному моделированию (scenario modeling), визуализации данных (data visualization), а также созданию и применению информационных панелей (dashboarding). Средства бизнес-аналитики сегодня находят применение во всех областях – от бюджетного планирования до расширенной аналитики (advanced analytics).

В традиционном понимании ведение DW относится только к структурированным данным (в этом

⁴⁴³ Большим данным и науке о данных в DMBOK2 посвящена отдельная глава, однако в качестве специальной области знаний по управлению данными, входящей в колесо DAMA (рис. 9.2), эта тема не выделена.

⁴⁴⁴ В главе 2 было приведено сравнение области бизнес-аналитики с «заводом по очистке данных» (рис. 2.7).

разделе основное внимание будет уделено вопросам построения и ведения DW именно в части таких данных). Однако с появлением новейших прогрессивных технологий к области BI и DW стали относить и управление полуструктурированными и неструктурированными данными (специфика BI/DW для этих данных рассматривается в разделе 14.3)⁴⁴⁵.

14.1.2. Цели и бизнес-драйверы

Внедряя у себя хранилища данных, организации преследуют следующие основные цели:

- поддержка деятельности в области BI;
- повышение эффективности бизнес-анализа и принятия решений;
- изыскание инновационных возможностей по результатам углубленного анализа данных.

Наиболее действенные драйверы развития хранилищ данных – необходимость сопровождения операционных функций, выполнения требований нормативно-правового соответствия и обеспечения деятельности в области бизнес-аналитики.

Однако главный драйвер – поддержка BI. Бизнес-аналитика нужна для полного понимания устройства и работы организации, ее клиентов и продуктов. Организация, деятельность которой основана на знаниях, полученных посредством грамотного бизнес-анализа, способна к неуклонному повышению эффективности и получению конкурентных преимуществ. По мере нарастания темпов поступления возрастающих объемов данных BI все более переходит от ретроспективной оценки к предиктивной аналитике.

Кроме того, в процессе операционной деятельности современной организации все чаще требуется наличие доказательных подтверждений соблюдения нормативно-правовых требований, подкрепленных историческими данными. Следовательно, системы управления хранилищами должны уметь обрабатывать и подобные запросы⁴⁴⁶.

14.1.3. Подходы к организации хранилища данных

Хранилища данных – это сравнительно новое технологическое решение, которое стало широко использоваться только в начале 1990-х годов, после того как Билл Инмон⁴⁴⁷, опубликовал в 1991 году свою первую книгу по этой теме – «Построение хранилища данных»⁴⁴⁸. Хотя отдельные элементы этой концепции и их технические воплощения существовали и ранее начиная с 1970-х годов, только к концу 80-х была в полной мере осознана необходимость интеграции корпоративной информации и надлежащего управления ею, а также появились технические возможности для создания соответствующих систем, первоначально названных хранилищами информации (information warehouse)⁴⁴⁹, а после выхода книги Инмона получивших свое нынешнее наименование хранилищ данных⁴⁵⁰.

На сегодняшний день существует два основных подхода к архитектуре хранилищ данных. Это так

445 DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

446 Там же.

447 Уильям «Билл» Инмон (William (Bill) Inmon) – известный американский ученый в области компьютерных наук, получивший всеобщее признание как «отец концепции хранилища данных».

448 В 2005 году вышло четвертое издание книги*.

* Inmon W. Building the Data Warehouse, 4th Edition. Wiley Publishing, Inc., 2005.

449 Devlin B. A., Murphy P. T. An Architecture for a Business and Information System // IBM Systems Journal, 1988. 27(1): 60–80. – URL: <http://altaplana.com/ibmsj2701G.pdf>.

450 Intersoft Lab. Сходство и различия двух подходов к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2057.phtml>.

называемая корпоративная информационная фабрика (Corporate Information Factory, CIF) Билла Инмона⁴⁵¹ и многомерное хранилище данных Ральфа Кимбалла^{452,453}.

Подход Инмона отражает метод проектирования «сверху вниз» и рассматривает хранилище как централизованное место хранения всех данных организации. После реализации централизованной модели данных для этого хранилища организации могут создавать на ее основе витрины данных (Data Marts, DM) – специальные хранилища для отдельных бизнес-направлений.

Подход Кимбалла основан на методе проектирования «снизу вверх». При этом подходе основным способом хранения данных являются витрины данных. Хранилище данных в целом представляет собой набор витрин, которые позволяют выполнять унифицированные аналитические задания, отчеты и другие необходимые процессы бизнес-аналитики⁴⁵⁴.

Рассмотрим эти подходы подробнее.

14.1.4. Корпоративная информационная фабрика (архитектура Инмона)

DW, согласно определению Инмона, представляет собой предметно-ориентированный, интегрированный, поддерживающий привязку ко времени, неизменяющийся набор сводных и детализированных исторических данных. Исходя из этого определения, можно выделить основные концептуальные компоненты, которые формируют отличия хранилища данных от операционных систем (систем поддержки операционной деятельности организации)^{455, 456}.

- **Предметная ориентированность:** данные в хранилище организованы по признаку соотнесения их с крупными сущностными объектами бизнеса, а не функциями или приложениями.

- **Интегрированность:** данные в хранилище унифицированы и связаны. Используются единообразные для всех компонентов хранилища структуры ключей, кодов шифрования, определений данных и условных наименований. Поскольку данные в хранилище интегрированы, они не являются простой копией операционных данных. Вместо этого DW, по сути, система записи (system of record) данных⁴⁵⁷:

- **Неизменяемость:** записи в DW обычно не обновляются, и этим хранилища принципиально отличаются от оперативных систем. Вместо обновления записи с новыми данными добавляются к уже имеющимся. А вот набор записей может отражать хронологию изменений состояния данных в процессе обработки одной и той же транзакции.

- **Привязка ко времени:** данные в записях DW сохраняются «как они есть» по состоянию на каждый заданный момент регистрации. По сути, записи в DW являются «моментальными снимками»

⁴⁵¹ Inmon W. Building the Data Warehouse, 4th Edition. Wiley Publishing, Inc., 2005.

⁴⁵² Ральф Кимбалл (Ralph Kimball) – известный американский специалист в области хранения данных и бизнес-аналитики, пропагандирующий методологию многомерного моделирования*.

* Kimball R. A Dimensional Modeling Manifesto. Kimball Group, 1997. – URL: <https://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto/>.

⁴⁵³ Kimball R., Margy R. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. John Wiley & Sons, Inc., 2013.

⁴⁵⁴ BI Consult. Современная архитектура хранилища данных: традиционное или облачное хранилище данных. BI Consult, 2021. – URL: <https://biconsult.ru/services/sovremennaya-arhitektura-hranilishcha-dannyh-tradicionnoe-ili-oblachnoe-hranilishche-dannyh>.

⁴⁵⁵ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁵⁶ Inmon W. Building the Data Warehouse, 4th Edition. Wiley Publishing, Inc., 2005.

⁴⁵⁷ Понятие «система записи» обсуждалось в главе 8.

состояния данных об описываемых объектах. Каждый снимок имеет метку времени. Как следствие, сколько бы вы ни запрашивали данные за один и тот же период времени, результаты выдачи будут неизменными вне зависимости от даты и времени обработки запроса.

• **Агрегированные и детализированные данные:** в DW сохраняются как записи о транзакциях на уровне мельчайших деталей, так и обобщенные данные. В операционных системах сводные данные обычно не учитываются. На заре создания DW необходимость обобщения данных диктовалась соображениями экономии вычислительных ресурсов и пространства памяти. В современных средах DW сводные данные могут иметься как на постоянном хранении (в табличной форме), так и формироваться по запросу (в режиме представления). Обычно решающим фактором при принятии решения о необходимости сохранения агрегированных таблиц является требуемая оперативность доступа к сводным данным.

• **Исторические данные:** операционные системы обрабатывают текущие данные, а в DW содержатся записи об истории операций, причем нередко в огромных объемах.

Хранилище на основе архитектуры Инмона построено в соответствии с реляционной моделью данных. Основные особенности реляционной модели были рассмотрены в главе 11 (см. раздел 11.2). Там мы, говоря о связях между сущностями и об атрибутах сущностей, выделили такие понятия, как первичный и внешний ключ. Прежде чем продолжить обсуждение особенностей корпоративной информационной фабрики, остановимся на понятии нормализации.

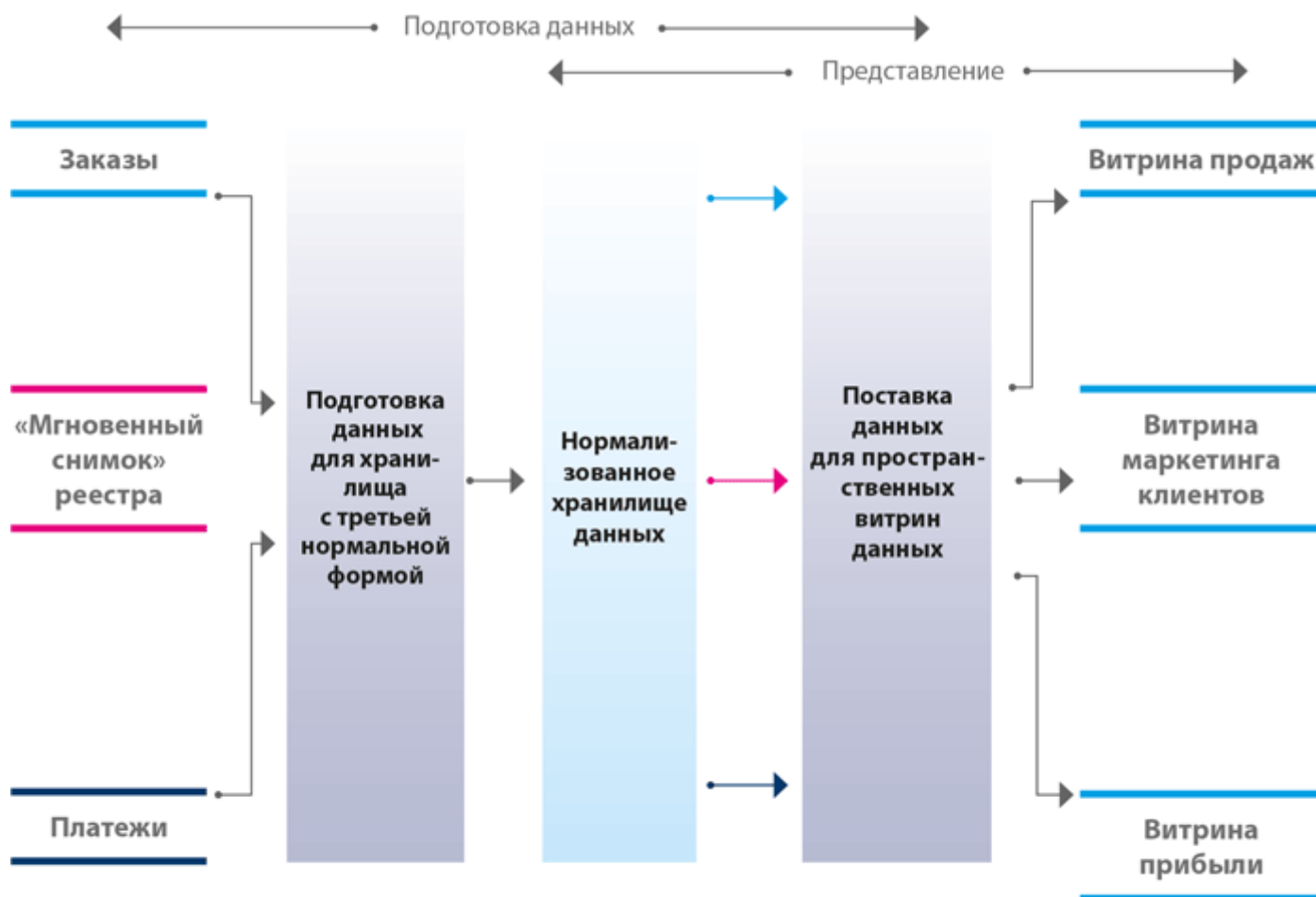


Рис. 14.1.

Корпоративная информационная фабрика*

* Intersoft Lab. Основные подходы к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2056.phtml>.

Нормализация (normalization) заключается в применении к модели данных наборов правил, позволяющих упорядочить необходимые для поддержки деятельности организации сведения в стабильные

структуры. Главная цель нормализации – сделать так, чтобы каждый атрибут содержался строго в одном месте во избежание избыточности и возможной противоречивости данных.

Правила нормализации разделяют и организуют атрибуты в соответствии с первичными и внешними ключами. Правила последовательно распределяются по уровням, и на каждом следующем уровне повышается степень детализации и добавляются новые требования по учету специфики сущностей при подборе корректных первичных и внешних ключей. Каждому уровню соответствует отдельная так называемая нормальная форма (normal form, NF). Всего выделяют пять нормальных форм (они обозначаются номерами в соответствии с уровнем), но на практике, как правило, достаточно третьей (3NF). Под нормализованной моделью обычно понимают данные, приведенные в форму 3NF⁴⁵⁸.

На рисунке 14.1 представлена укрупненная архитектура корпоративной информационной фабрики.

Подготовка данных начинается со скоординированного извлечения их из источников. После этого осуществляется загрузка реляционной базы данных, которая в итоге содержит детализированные (атомарные) данные в третьей нормальной форме. Наполненное нормализованное хранилище используется для того, чтобы снабжать информацией дополнительные репозитории презентационных данных (данных, подготовленных для анализа). Эти репозитории, в частности, включают специализированные хранилища для изучения и извлечения информации (data mining), а также витрины данных.

С целью представления создаются отдельные витрины агрегированных данных, предназначенные для обслуживания бизнес-подразделений или для реализации бизнес-функций. Для структурирования данных в них используется многомерная модель (о ней мы поговорим в следующем разделе). Детализированные данные при этом остаются доступными, что обеспечивается с помощью нормализованного хранилища. Таким образом, структура детализированных и агрегированных данных при такой архитектуре существенно различается.

Подводя итог вышесказанному, можно выделить следующие отличительные характеристики архитектурного подхода Инмона⁴⁵⁹.

- Использование реляционной модели организации детализированных данных и многомерной – для организации агрегированных данных.
- Использование итеративного подхода при создании больших хранилищ данных, построение хранилища не сразу, а по частям. Это позволяет при необходимости вносить изменения в небольшие блоки данных или программных кодов и избавляет от необходимости реструктурировать значительные объемы данных или осуществлять сложное перепрограммирование. То же можно сказать и о потенциальных ошибках: они также будут локализованы в пределах сравнительно небольшого массива данных без риска испортить все хранилище.
- Использование третьей нормальной формы для организации детализированных данных обеспечит высокую степень детализации интегрированных данных и предоставит организации широкие возможности для манипулирования, изменения формата и способа представления данных по мере необходимости.
- Хранилище данных – это проект корпоративного масштаба, охватывающий все подразделения и обслуживающий нужды всех пользователей корпорации.
- Хранилище данных – это не механическая коллекция витрин данных, а физически целостный объект.

14.1.5. Многомерное хранилище данных (архитектура Кимбалла)

Архитектурный подход к представлению хранилища данных Кимбалла основан на многомерной структурной модели данных. Кимбалл определяет хранилище как копию транзакционных данных, особым образом структурированную для обработки запросов и анализа. «Копия» в данном контексте не означает точной копии оригинала. При переносе в хранилище данные подвергаются реструктуризации для обеспечения соответствия схеме многомерной модели, которая специально проектируется таким образом,

⁴⁵⁸ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁵⁹ Intersoft Lab. Основные подходы к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2056.phtml>.

чтобы сделать данные предельно понятными и полезными для потребителей, но при этом сохранить и достаточный для обработки запросов уровень формализации. Важнейшее отличие многомерных схем хранения данных от традиционных реляционных – отказ от нормализации^{460, 461}.

Многомерные модели, часто называемые также звездообразными схемами (star schema), представляют собой подборки фактов (facts), под которыми понимаются числовые данные или характеристики бизнес-процессов (например, объем продаж) в проекции на измерения (dimensions), которые используются для описания атрибутов, соответствующих фактам и позволяющих пользователям правильно интерпретировать фактические данные (например, с объемом продаж сопоставляются артикул продукта X и отчетный квартал). Таблица фактов связана со множественными таблицами измерений, и в графическом представлении такая схема организации данных имеет форму звезды, откуда и возникло название. При наличии в модели множественных таблиц фактов они проецируются на общие для различных таблиц так называемые конформные (conformed) измерения через шину (bus), подобную компьютерной шине⁴⁶². Множественные витрины данных на корпоративном уровне могут интегрироваться посредством подключения их к общей шине конформных измерений.

Матрица шины DW отражает доступные фактические данные на пересечениях строк бизнес-процессов (фактов) и столбцов предметных областей (измерений). Возможности для интеграции через конформные измерения появляются там, где множественные процессы используют одни и те же данные⁴⁶³.

В таблице 14.1 приведен простейший пример матрицы шины DW.

К бизнес-процессам отнесены продажи, запасы и заказы. Данные обо всех трех бизнес-процессах могут интегрироваться через общие для них конформные измерения Дата и Продукт.

Данные о продажах и запасах могут интегрироваться через измерение Магазин, а данные о запасах и заказах – через измерение Поставщик. Таким образом, лишь четыре измерения из пяти – Дата, Продукт, Магазин и Поставщик – являются кандидатами на роль конформных. А вот измерение Склад общим для каких-либо бизнес-процессов не является и для интеграции данных непригодно, поскольку ему соответствует единственный бизнес-процесс – учет запасов.

⁴⁶⁰ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁶¹ Kimball R., Margy R. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. John Wiley & Sons, Inc., 2013.

⁴⁶² Термин «шина» Кимбалл позаимствовал из своей первоначальной области специализации – электротехнической инженерии, где к общей шине питания подключаются различные энергопотребляющие компоненты системы*.

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁶³ Там же.

Таблица 14.1.

Пример матрицы шины DW организации*

	Предметные области				
	Дата	Продукт	Магазин	Поставщик	Склад
Бизнес-процессы					
Продажи	X	X	X		
Запасы	X	X	X	X	X
Заказы	X	X		X	
Потенциальная конформность измерения	Да	Да	Да	Да	Нет

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

На рисунке 14.2 представлена укрупненная архитектура многомерного хранилища данных.

**Рис. 14.2.**

Многомерное хранилище данных*

* Intersoft Lab. Основные подходы к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM

World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2056.phtml>.

Как и при архитектурном подходе Инмона, подготовка данных начинается со скоординированного извлечения их из источников. При этом уже на этапе подготовки первичные данные преобразуются в вид, пригодный для использования (с учетом требований к скорости обработки информации и качеству данных). Ряд операций совершается централизованно, например поддержание и хранение общих справочных данных, а другие операции могут выполняться распределенно.

В хранилище (области представления) содержатся такие же детализированные данные, как и в нормализованном хранилище Инмона, однако они структурированы в соответствии с многомерной моделью (что облегчает использование данных и выполнение запросов). При этом хранилище может быть централизованным или распределенным.

Хранилище содержит как детализированные, так и агрегированные данные, сформированные в соответствии требованиями в части производительности или пространственного распределения.

Запросы в процессе выполнения могут оперировать на различных уровнях детализации без дополнительного перепрограммирования со стороны пользователей или разработчиков приложений.

В отличие от архитектуры CIF, многомерные модели строятся для обслуживания бизнес-процессов (которые в свою очередь связаны с бизнес-показателями или бизнес-событиями), а не бизнес-подразделений. Например, данные о заказах, которые должны быть доступны для общекорпоративного использования, вносятся в многомерное хранилище только один раз (при подходе Инмона их пришлось бы трижды копировать в витрины данных отделов маркетинга, продаж и финансов). После формирования в хранилище сведений об основных бизнес-процессах консолидированные модели могут обеспечивать выдачу их кросс-процессных характеристик. С развитием матрицы корпоративного хранилища данных с архитектурой шины происходит расширение связей между показателями бизнес-процессов (фактами) и описательными атрибутами (измерениями)⁴⁶⁴.

На основе вышесказанного можно выделить следующие отличительные характеристики архитектурного подхода Кимбалла⁴⁶⁵:

- использование многомерной модели данных;
- хранилище включает как детализированные данные о транзакциях, так и агрегированные.
- хранилище данных не служит единым физическим репозиторием (в отличие от подхода Инмона).

Это виртуальное хранилище, представляющее собой набор витрин данных, каждая из которых имеет архитектуру «звезда».

14.1.6. Сравнение архитектурных подходов Инмона и Кимбалла

Сходства

Прежде всего, и тот и другой подход направлены на создание одного объекта – корпоративного хранилища данных. Единство конечного объекта означает общность требований, которым должен удовлетворять любой подход для достижения искомого конечного результата, а это, в свою очередь, указывает на то, что и в самой архитектуре должны быть общие черты.

Первое основное требование связано с тем, что для принятия и осуществления важных решений все организации нуждаются в средстве для обеспечения хранения, анализа и интерпретации данных, которые они накапливают. Именно для этого создаются хранилища данных.

Второе требование – это требование точности и своевременности предоставления данных. Каждый пользователь должен иметь возможность доступа к любым данным в соответствии со своими конкретными требованиями, и этот доступ должен осуществляться с помощью понятных и простых способов построения запросов.

Оба архитектурных подхода отвечают названным требованиям в полной мере. Они обеспечивают хранение как детализированных, так и агрегированных данных, которые доступны пользователям для

⁴⁶⁴ Intersoft Lab. Основные подходы к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2056.phtml>.

⁴⁶⁵ Там же.

анализа⁴⁶⁶.

Отличия

Первое существенное отличие между этими архитектурами – различные подходы к построению баз данных, составляющих основу хранилища. Если в архитектуре Кимбалла используется многомерная модель данных как на стадии подготовки, так и презентации данных, то архитектура Инмона комбинирует два подхода. В СІF детализированные данные организованы в виде нормализованной реляционной базы, а предназначенные для представления пользователям агрегированные данные организованы с помощью многомерных моделей, как и у Ральфа Кимбалла.

Таким образом, архитектуры отличаются, по сути, только способом организации детализированных данных.

Второе принципиальное отличие этих двух подходов, отчасти вытекающее из первого, – вопрос физической организации хранилища. Если у Инмона хранилище данных – это физически целостный реально существующий объект, то хранилище Кимбалла – скорее «виртуальный» объект. Это набор витрин данных, которые могут быть пространственно разобщенными⁴⁶⁷.

Преимущества и недостатки

Преимущества и недостатки каждого из подходов напрямую вытекают из применяемых ими моделей данных.

С одной стороны, многомерная модель, на которую ориентирована архитектура Кимбалла, облегчает доступ к данным и требует меньше времени на выполнение запросов, а также упрощает работу с детализированными данными. С другой стороны, такая организация данных критикуется за отсутствие необходимой гибкости и чувствительности структуры к изменениям, поскольку в детализированные данные, представленные многомерной моделью, труднее вносить необходимые корректировки.

Реляционная модель детализированных данных, используемая в хранилище Инмона, замедляет доступ к данным и требует больше времени для выполнения запросов в силу различной организации детализированных и агрегированных данных. С другой стороны, она предоставляет широкие возможности для оперирования детализированными данными и изменения их формата и способа представления по мере необходимости.

В целом выбор того или иного архитектурного решения определяется нуждами бизнеса и его конкретными особенностями⁴⁶⁸.

⁴⁶⁶ Intersoft Lab. Сходство и различия двух подходов к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2057.phtml>.

⁴⁶⁷ Там же.

⁴⁶⁸ Там же.

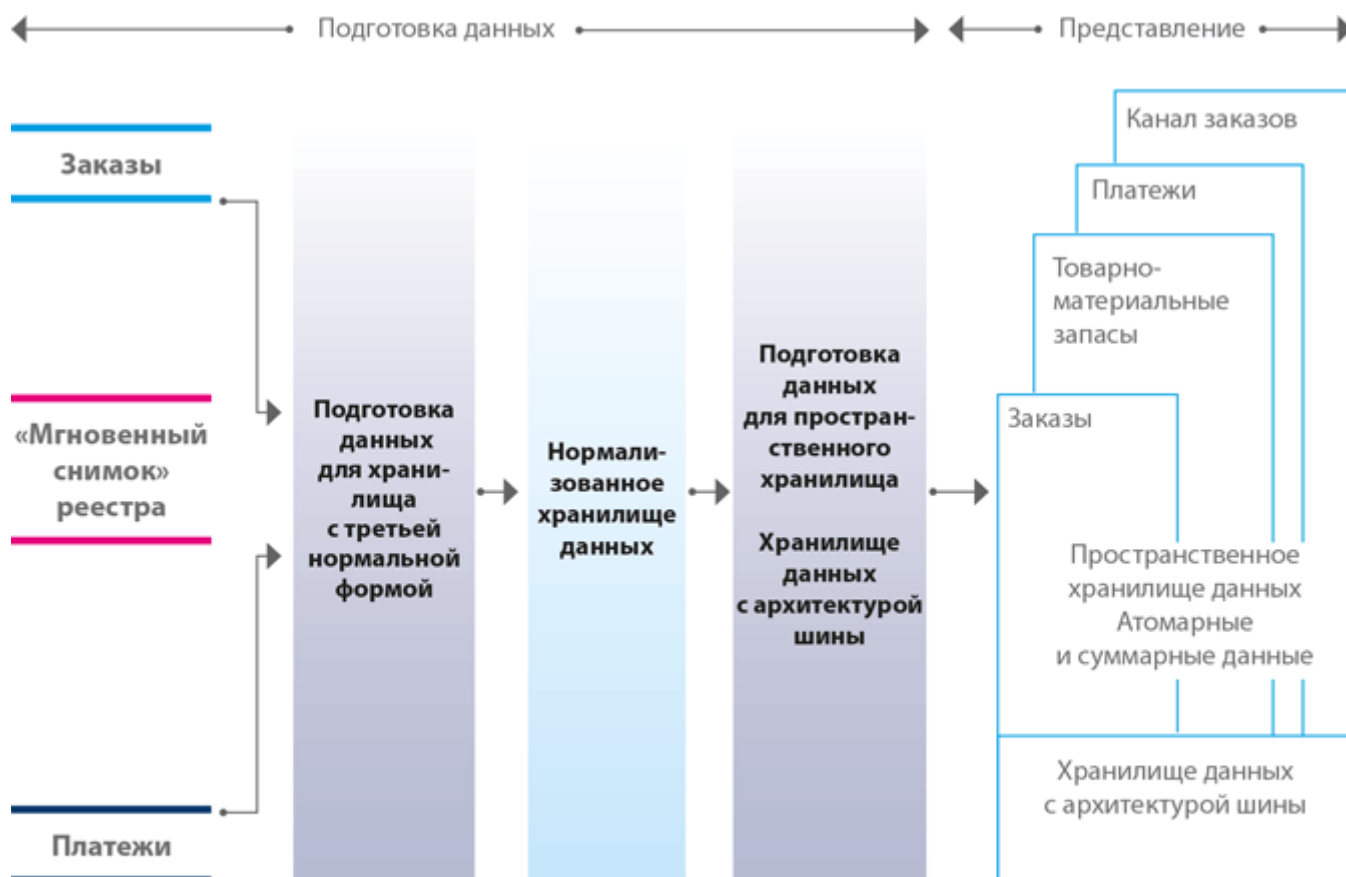


Рис. 14.3.

Гибридная архитектура хранилища данных*

* Intersoft Lab. Сходство и различия двух подходов к архитектуре Хранилищ данных. Intersoft Lab: Журнал BPM World. 2005. – URL: <https://iso.ru/ru/press-center/journal/2057.phtml>.

14.1.7. Гибридный подход

Многие организации используют так называемый гибридный подход, стараясь совместить преимущества обоих описанных выше архитектурных концепций.

Гибридная архитектура хранилища данных представлена на рисунке 14.3.

Гибридное хранилище включает нормализованное хранилище CIF и многомерное хранилище детализированных и агрегированных данных на основе архитектуры шины.

Недостаток данного подхода в том, что двойная работа по подготовке и хранению детализированных данных сопровождается существенными дополнительными расходами и задержками⁴⁶⁹.

14.1.8. Обобщенная архитектура аналитической рабочей среды организации

С развитием концепции больших данных представление об архитектуре аналитической рабочей среды организации несколько меняется, поскольку контур обработки больших данных стал дополнительным магистральным каналом притока новых сведений. Рисунок 14.4 описывает обобщенную архитектуру рабочих сред организации для областей бизнес-аналитики (на основе традиционного хранилища данных – DW) и науки о данных (на основе хранилища больших данных).

Работа с данными в среде DW/BI осуществляется следующим образом. Из систем-источников данные поступают в область временного хранения, где подвергаются очистке и обогащению, интегрируются и отправляются на хранение в центральное хранилище данных (DW) или хранилище операционных данных (operational data store, ODS). Доступ к данным из DW осуществляется через витрины или кубы (data cube).

⁴⁶⁹ Там же.

Хранилище операционных данных представляет собой интегрированную базу операционных данных, поступающих от приложений или из других баз данных. В ODS обычно содержатся только текущие данные или данные за относительно небольшой отчетный период, в то время как в главном DW накапливаются еще и исторические данные. Главное же отличие ODS от DW заключается в том, что операционные данные динамически изменяются по мере поступления новых данных в отличие от статичных данных в главном хранилище. ODS используются далеко не во всех организациях, а только в тех, где требуется минимизировать время запаздывания.

Для проведения бизнес-анализа в настоящее время предлагается широкий спектр BI-инструментов, которые можно разбить на следующие основные типы⁴⁷⁰:

- операционная отчетность – позволяет выявлять и анализировать краткосрочные (помесячные) и среднесрочные (годовые) тенденции и закономерности;
- управление эффективностью бизнеса (Business Performance Management, BPM) – позволяет производить формальную оценку измеримых показателей, соответствующих целям организации;
- приложения для оперативного анализа – могут включать функции анализа клиентов, финансов, цепочек поставок, организации производства, управления персоналом и т. п.

Среди приложений для оперативного анализа особо выделяются инструменты онлайн-аналитической обработки (online analytical processing, OLAP), обеспечивающие высокопроизводительную обработку многомерных аналитических запросов. Выдача данных в ответ на запросы обычно происходит в матричном формате. Измерения определяются столбцами и строками матрицы, на пересечении которых выводятся факторы или значения. Концептуально такая модель может быть представлена как многомерный куб данных. Ниже перечислены три наиболее распространенные архитектуры OLAP-систем.

- Реляционная (ROLAP): функциональность OLAP реализуется посредством моделирования многомерности через определение связей между атрибутами стандартных двумерных таблиц систем управления реляционными базами данных. Стандартная схема модели данных в среде ROLAP – звездообразная.
- Многомерная (MOLAP): поддержка OLAP в составе или с использованием коммерческих и специализированных многомерных баз данных.
- Гибридная (HOLAP): сочетание ROLAP и MOLAP. Гибридные реализации позволяют хранить часть данных в MOLAP, а часть – в ROLAP. Реализации могут варьироваться в зависимости от имеющихся у проектировщика возможностей по контролю структуры разделов данных.

Важнейший принцип организации портфеля BI-приложений – самообслуживание (self-service) в части настроек представления и выдачи данных. Доступные пользователю действия обычно регулируются настройками профиля на портале доступа, где, в зависимости от привилегий, можно выбирать, подключать или отключать и конфигурировать различные функциональности, уведомления, сообщения и предупреждения, периодичность просмотра производственных отчетов, порядок взаимодействия с аналитическими отчетами, разрабатывать собственные отчеты и пользоваться настройками и функциями приборной панели и картами показателей.

Выделенные на рисунке 14.4 архитектурные компоненты в различных организациях могут быть реализованы по-разному, в зависимости от выбранного архитектурного подхода. В частности, как мы уже говорили, архитектура Кимбалла подразумевает, что данные в хранилище структурно распределены по витринам данных подразделений, с помощью которых и обеспечивается их очистка, стандартизация и управление. В этом случае именно в витринах хранится история данных на максимально детализированном (атомарном) уровне⁴⁷¹.

Параллельно обработке данных в среде DW/BI во многих организациях осуществляется обработка входящих потоков больших данных. При этом данные сначала загружаются в специальное хранилище – озеро данных (data lake), а затем осуществляется их интеграция и исследование с построением моделей. Работа с большими данными будет рассмотрена более подробно далее (см. раздел 14.3).

⁴⁷⁰ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁴⁷¹ Там же.

14.1.9. Контекстная диаграмма области знаний и уровни зрелости функции «Ведение хранилищ данных и бизнес-аналитика»

Контекстная диаграмма области знаний «Ведение хранилищ данных и бизнес-аналитика» представлена на рисунке 14.5.

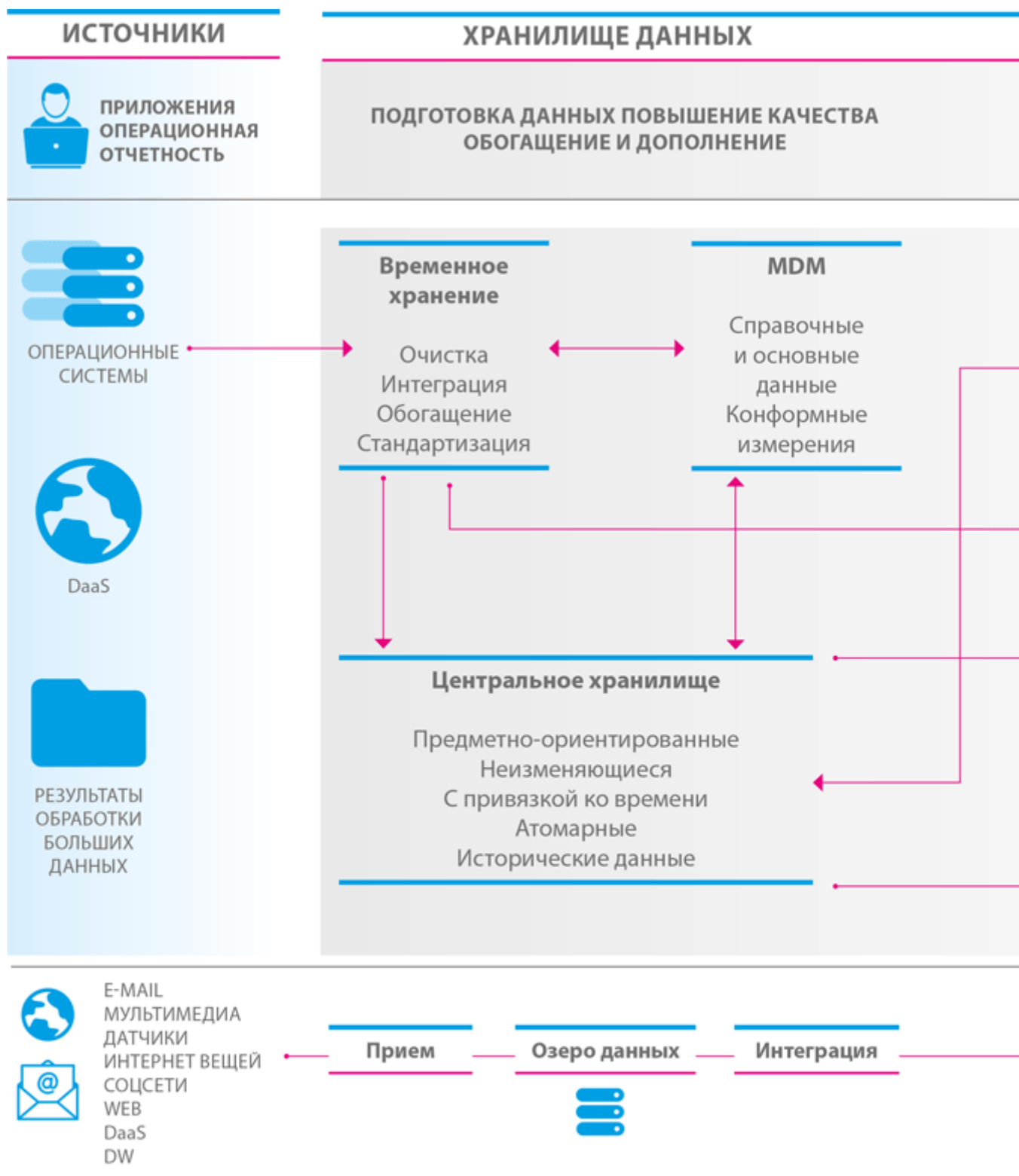
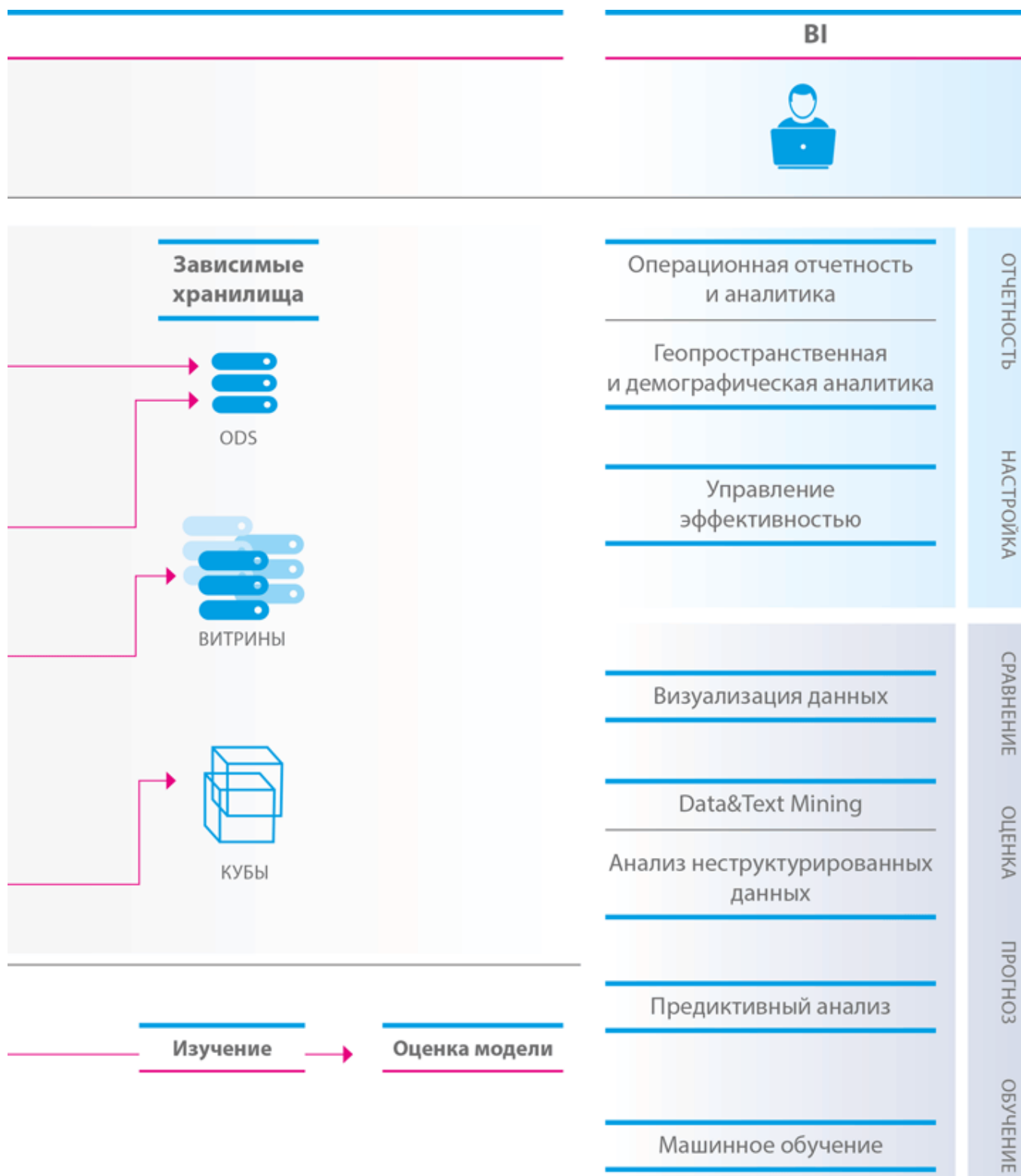


Рис. 14.4.

Обобщенная архитектура комплексной рабочей среды для областей DW/BI и работы с большими данными*



* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

При планировании и внедрении хранилища данных следует руководствоваться следующими принципами:

- фокусировка на целях бизнеса. Хранилище данных должно соответствовать приоритетам организации и способствовать решению бизнес-задач;

- ориентация на желаемые конечные результаты. Приоритеты и интересы бизнеса плюс потребности в данных BI-приложений должны от начала и до конца диктовать выбор содержания и структуры информационного наполнения хранилища данных;

- мыслить глобальными категориями при планировании архитектуры, но руководствоваться локальными соображениями при построении. Видение полной картины конечного результата воплощения архитектурного замысла – требование обязательное, но реализация этого замысла ведется итерационно-поступательными движениями – целевыми проектами или «спринтерскими рывками», обеспечивающими быструю окупаемость вложений;

- обобщение и оптимизация производятся на завершающих, а не начальных этапах реализации. Системную архитектуру необходимо выстраивать на основе максимально детализированных данных. Обобщение, сведение и интеграцию с целью приведения структуры данных к стандартным требованиям и повышения производительности систем нужно на время отложить, поскольку для восстановления утерянных деталей всю работу придется откатывать до точки дезинтеграции;

- ориентация на прозрачность и самообслуживание. Чем больше контекста (т. е. всевозможных метаданных), тем проще потребителям разобраться в смысле данных и найти им полезное и выгодное применение. Необходимо информировать заинтересованные стороны о происхождении данных и процессах их интеграции;

- параллельно с хранилищем необходимо выстраивать метаданные. Критический фактор успеха хранилища данных – способность объяснять смысл и происхождение данных. Структура метаданных должна формироваться на стадии проработки модели данных, а учет и управление – входить в состав рабочих процессов и текущих операций⁴⁷²;

- сотрудничество с другими направлениями и проектами в области управления данными. Прежде всего необходимо осуществлять тесное взаимодействие с ответственными за руководство данными, обеспечением качества данных и ведением метаданных;

- нельзя подходить ко всем потребителям данных с едиными критериями. Различным группам потребителей требуются различные инструменты и продукты⁴⁷³.

В начале раздела 14.1 мы уже отметили роли DW и BI в цепочке поставок данных. Внедренное хранилище данных и его ориентированная на потребителей данных часть, включающая клиентские приложения и инструменты BI, превращаются, по сути, в информационный продукт. Последующие усовершенствования платформы DW (дополнения, надстройки и/или модификации) следует проводить поэтапно, методом приращений.

472 Управление метаданными мы будем обсуждать в следующей главе, посвященной вопросам базовой поддержки жизненного цикла данных.

473 Там же.

ВЕДЕНИЕ ХРАНИЛИЩ ДАННЫХ И БИЗНЕС-АНАЛИТИКА

Определение: планирование, внедрение и контроль процессов предоставления данных для принятия решений и информационная поддержка специалистов, участвующих в аналитической деятельности и формировании отчетности

Цели:

1. Создание и сопровождение ИТ-среды, технических и бизнес-процессов, необходимых для обеспечения интегрированными данными деятельности по выполнению операционных функций, поддержке нормативно-правового соответствия и проведению бизнес-анализа.
2. Поддержка и обеспечение эффективного бизнес-анализа и принятия решений.

БИЗНЕС-ДРАЙВЕРЫ

Входные материалы	Проводимые работы	Результаты
• Бизнес-требования • Требования по масштабируемости, проведению операций, инфраструктуре, а также технической поддержке • Требования по качеству, безопасности и доступу к данным • ИТ-стратегия • ИТ-политики и стандарты • Внутренние источники данных • Справочные и основные данные • Отраслевые и внешние данные	1. Выработка понимания требований (П) 2. Определение и сопровождение архитектуры DW/BI (П) 3. Проектирование и разработка хранилища а и витрин данных (Р) 4. Заполнение хранилища данных (Р) 5. Внедрение портфеля инструментов BI (Р) 6. Сопровождение информационных продуктов (О)	• Архитектура DW/BI • Информационные продукты • Процесс заполнения • Деятельность по руководству • Описание происхождения • План обучения и перехода к применению на практике • План выпуска релизов • Процесс сопровождения информационного продукта • Меры по оптимизации нагрузки • Мониторинг использования средств BI
Поставщики	Участники	Потребители
• Руководители направлений бизнеса • Совет по руководству данными • Корпоративные архитекторы • Производители данных • Потребители информации • Эксперты в предметных областях	• Спонсоры и владельцы продукта • Архитекторы и аналитики • Специалисты в области DW/BI (платформа BI, хранилища, управление информацией) • Руководители проектов • Ответственные за управление изменениями	• Потребители информации • Клиенты • Руководители и менеджеры

ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы	Инструменты	Метрики
• Прототипирование с целью уточнения требований • Бизнес-аналитика самообслуживания • Данные аудита, доступные по запросам	• Репозитории метаданных • Средства интеграции данных • Аналитические приложения	• Метрики использования • Удовлетворенность потребителей/пользователей • Востребованность данных по предметным областям (%) • Время ответа / производительность

(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 14.5.

Контекстная диаграмма. Ведение хранилищ данных и бизнес-аналитика*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

На рисунке 14.6 представлены обобщенные характеристики уровней зрелости функции «Ведение хранилищ данных и бизнес-аналитика».

Практики в области ведения хранилищ данных и бизнес-аналитики составляют основную часть стратегии организации, а также рассматриваются как существенный фактор для каждого проекта тактического уровня. Результаты обратной связи и оценка предыдущего опыта используются для выработки лучших общеорганизационных практик. Проявляется активное стремление к дальнейшему совершенствованию.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Практики в области ведения хранилищ данных и бизнес-аналитики измеряются и контролируются в соответствии с общеорганизационными стандартами. Все подразделения отчитываются о соответствии или несоответствии совету по руководству данными, используя KPI. Контроль охватывает деятельность по предоставлению централизованно управляемого хранилища данных, а также возможностей в части BI для обеспечения потребностей подразделений в отчетности и аналитике. Осуществляется поддержка со стороны центра компетенций в сфере бизнес-аналитики.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Для поддержки стратегической линии в области ведения хранилищ данных и бизнес-аналитики во всех подразделениях и функциональных направлениях определены общеорганизационные политики и стандарты. Подразделения выстраивают локальные практики предоставления отчетности на основе этих стандартов.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

В конкретных подразделениях или отдельных проектах успешно внедряются, документируются и повторно применяются практики бизнес-анализа. Могут существовать локальные политики и стандарты, но нет никаких свидетельств того, что они координируются на более высоком уровне. Это приводит к применению несовместимых подходов или дублированию усилий в рамках всей организации.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации отсутствуют какие-либо свидетельства наличия последовательной стратегии в области бизнес-аналитики, поскольку существуют значительные различия в том, как различные бизнес-подразделения обрабатывают информацию. Любая отчетность или информационные панели, как правило, тесно привязаны к конкретным программным продуктам и сервисам.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 14.6

Обобщенные характеристики уровней зрелости функции «Ведение хранилищ данных и бизнес-аналитика»*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

14.1.10. Влияние на ценность данных

Довольно часто встречаются организации, у которых есть несколько хранилищ данных. Например, отдельные хранилища для подразделений финансов, продаж и маркетинга, обслуживания клиентов и т. п. Нередко одни и те же системы поддержки операционной деятельности загружают данные в несколько

хранилищ, а иногда данные передаются между этими хранилищами.

Однако усилия по интеграции, стандартизации и обработке данных в хранилище обеспечивают наибольшую отдачу, если данные становятся доступными для всех, кто ими пользуется. Таким образом, максимальный эффект от внедрения технологий DW достигается только при создании корпоративного хранилища данных (EDW), собирающего данные от всех операционных систем и делающего их доступными для всей организации.

Можно выделить следующие преимущества внедрения в организации EDW, которые оказывают существенное влияние на повышение ценности ее данных⁴⁷⁴.

Доступность данных для всей организации

Прежде чем кто-то в организации попытается инициировать реализацию нового отчета, процедуры по извлечению, преобразованию и загрузке (ETL), сервиса данных и т. п., он должен ознакомиться с возможностями, предоставляемыми EDW. Может оказаться, что нечто подобное уже доступно где-то в организации.

Необходимое условие обеспечения доступности данных – налаженное должным образом ведение и обеспечение доступности соответствующих метаданных. Все заинтересованные сотрудники организации должны знать, какие данные имеются в ее распоряжении (см. главу 15).

Возможность публикации данных в режиме, близком к реальному времени

В главе 11, обсуждая задержки при обработке данных, мы говорили о том, что обработка данных о текущих операциях обычно проводится в режиме реального времени или в режиме, близком к реальному времени (near real-time), а данных, требуемых для анализа и отчетности, – по графику, в пакетном режиме (см. главу 12, раздел 12.2.3). Однако с учетом постоянного ужесточения бизнес-требований пакетный режим во многих случаях становится все менее приемлемым.

Современные технологии позволяют обнаруживать и передавать изменения в данных почти в режиме реального времени (например, на основе модели публикации и подписки – см. главу 12, раздел 12.2.3).

Публикация данных в режиме, близком к реальному времени, не всегда проста и в значительной степени зависит от операционных систем, предоставляющих данные в хранилище данных. Поэтому при реализации таких возможностей важно обеспечить их доступность на корпоративном уровне максимально широкому кругу сотрудников.

Возможность отслеживания происхождения данных

Завоевание доверия к информации – важнейший аспект внедрения DW. По мере поступления данных в хранилище вместе с ними должны передаваться сведения о том, откуда они появились. Это достигается за счет описания происхождения данных (data lineage) (см. главы 8 и 11) и организации доступности соответствующих метаданных (см. главу 15). Понимая, кому принадлежат отдельные элементы данных и для каких целей они были получены, мы можем добиться соответствия нормативно-правовым требованиям и использовать данные эффективнее.

Роль EDW как корпоративной памяти

Хранилище данных должно выступать в качестве корпоративной памяти. Под этим подразумевается, что оно должно хранить с привязкой ко времени все данные, которые использует организация.

Когда руководство организации принимает решение, подкрепленное какими-либо доказательствами, хранилище данных и связанные с ним инструменты бизнес-аналитики должны обеспечивать предоставление исторической справки об этих доказательствах. Это не только усиливает поддержку выполнения требований нормативно-правового соответствия, но и позволяет в будущем развивать анализ тенденций.

Кроме того, существенная помощь обеспечивается с точки зрения извлечения уроков. Сравнение фактических результатов решения с результатами, предсказанными на основе данных, может быть

⁴⁷⁴ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

использовано для повышения эффективности в будущем.

Эффективное использование опыта экспертов

Как правило, во многих подразделениях организации есть сотрудники – «источники знаний», которые хорошо разбираются в тонкостях информационных систем и данных. Объединив этих людей с теми, кто понимает, как применять инструменты анализа, можно получить много новых идей в отношении данных и их использования.

Целесообразно создать в организации центр компетенций в сфере бизнес-аналитики. Он должен играть роль центрального источника знаний для всех, у кого есть вопросы или предложения по поводу использования EDW. Деятельность центра также будет стимулировать инновации (что положительно скажется на итоговой прибыли) и может даже повысить привлекательность организации для потенциальных сотрудников.

14.2. Управление документами и контентом

Управление документами и контентом подразумевает наличие средств, позволяющих контролировать создание, регистрацию, хранение, защиту и использование самых разнородных сведений, доступ к которым невозможно организовать с помощью традиционных реляционных систем управления базами данных. Процессы в цепочках поставок таких данных имеют свою специфику, которую мы обсудим в этом разделе.

14.2.1. Определение области знаний «Управление документами и контентом»

Управление документами и контентом (Document and Content Management, DCM) распространяется на любые неструктурированные и полуструктурированные данные, т. е. данные, структура которых не описывается предопределенной моделью – будь то реляционная или любая другая модель данных (см. главу 8, раздел 8.2). Главная задача здесь – обеспечение сохранности и целостности обрабатываемых сведений, а также регулирование доступа к ним. В этом плане управление документами и контентом в общих чертах мало чем отличается от операционного управления реляционными базами данных (см. главу 12, раздел 12.1). Однако помимо текущих задач в этой области есть и стратегические. Во многих организациях неструктурированные данные имеют прямое отношение к структурированным. Соответственно, и управленческие решения должны приниматься согласованно и применяться последовательно. Кроме того, документы и неструктурированный контент требуют надежной защиты и контроля качества. А обеспечение информационной безопасности и качества таких данных невозможно без руководства, надежной архитектуры и налаженного управления метаданными⁴⁷⁵.

14.2.2. Цели и бизнес-драйверы

Основными целями управления данными и контентом являются:

- обеспечение возможностей для эффективного накопления, получения и использования данных, сохраняемых в неструктурированных форматах;

- интеграция структурированных и неструктурированных данных;

- соблюдение действующего законодательства и обеспечение соответствия ожиданиям клиентов.

Главными бизнес-драйверами управления документами и контентом являются обеспечение соблюдения требований нормативно-правового регулирования, способность адекватно отвечать на запросы судебно-арбитражных и надзорных органов об электронном раскрытии информации (e-discovery), а также обеспечение непрерывности бизнеса. Составители DMBOK считают, что качественное управление рассматриваемым видом данных оказывает серьезное влияние на эффективность работы организаций. Прогресс в области технологий управления документами позволяет организациям автоматизировать и оптимизировать документооборот, устранять дублирующие друг друга ручные операции, нормализовать внутриорганизационное сотрудничество и внешние партнерские связи.

⁴⁷⁵ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

В свою очередь, дополнительным преимуществом этих технологий выступает упрощение и ускорение поиска, доступа и публикации нужных документов. Наконец, они же способствуют предотвращению утери важных документов. Все это крайне важно и для обеспечения электронного раскрытия информации, и для экономии денежных средств за счет высвобождения офисного пространства и снижения затрат на ведение и обработку документации.

Хорошо структурированные веб-сайты с мощными поисковыми возможностями позволяют эффективно управлять онтологиями и другими структурами, максимально упрощающими клиентам и сотрудникам поиск нужного контента, тем самым повышая уровень их удовлетворенности⁴⁷⁶.

14.2.3. Основные понятия в области управления документами и контентом

Говоря об управлении документами и контентом, следует пояснить ряд исходных понятий⁴⁷⁷.

Документы и записи

Документами называются электронные или бумажные (как печатные, так и рукописные) материалы с инструкциями, руководствами, требованиями и распоряжениями, которые касаются выполнения различных задач или функций, например, протоколами собраний и решений. Также документы могут использоваться для распространения информации или обмена знаниями и опытом. Примеры распространенных типов документов – акты, методики, правила, протоколы, процедуры руководства, спецификации, стандарты и технические задания.

В свою очередь, записи (records) – это подмножество документов определенного вида. То есть не всякий документ классифицируется как запись, но всякая запись относится к категории документов. Записи свидетельствуют, что действия, сведения о которых в них зафиксированы, были действительно произведены и сделано это было в установленном нормативными документами порядке; соответственно, записи могут использоваться для представления, например, в надзорные органы в качестве доказательства соблюдения организацией в ходе осуществления текущей деятельности установленных всевозможными регламентами требований. В современных условиях записи создаются или ведутся не только людьми, но и автоматическими средствами мониторинга и регистрации.

Документы и записи бывают физическими (документы, записки, договора, отчеты, квитанции, письма и микрофильмы) и электронными (письма и вложенные файлы e-mail, СМС, сообщения по мессенджерам и многие другие.). Отдельно стоит выделить контент веб-сайтов, документы на любых носителях и в любых аппаратных средах, а также записи в базах данных любого рода и типа.

Контент

Контент – это информационное наполнение документа. Документ можно рассматривать как некий контейнер, а вот контент – как то, что в этом контейнере содержится. Под контентом принято понимать данные и информацию, размещенную внутри файла, документа или на веб-сайте. Контентом часто управляют исходя из степени концептуальной важности документов, в которых он содержится, а также в зависимости от типа или статуса документов. Важно отметить, что у контента есть свой вполне конкретный жизненный цикл. В своей завершенной форме часть контента становится содержимым записей организации. Официальные записи требуют особого обращения по сравнению с прочим контентом.

Управление документами и записями

Управление документами (document management) – понятие, которое описывает весь спектр процессов, приемов и технологий распоряжения документами и записями на протяжении всего их жизненного цикла, включая хранение, учет и контроль как электронных, так и бумажных вариантов. Свыше 90 % документов в наши дни создаются в электронной форме.

В целом управление документами занимается их формой, а не содержанием, – иными словами, файлами и папками, а не контентом. Информационное наполнение может служить лишь подсказкой, как

476 Там же.

477 Там же.

лучше этим файлом распоряжаться, но в рамках управления документами практическое обращение с этим файлом как с документом будет все так же строиться на основе его рассмотрения как единого и неделимого целого.

Управление записями (records management) – важнейший компонент управления документами. К управлению записями предъявляются особые требования. Согласно международному стандарту ISO 15489–1:2016 Information and documentation – Records management – Part 1: Concepts and principles управление записями включает⁴⁷⁸:

- а) создание записей и их ввод в систему с целью доказательства ведения деловых операций;
- б) принятие надлежащих мер по защите их аутентичности, достоверности, целостности и пригодности для использования в условиях изменяющейся во времени деловой среды и требований к управлению записями.

Управление жизненным циклом документов и записей предполагает следующие виды работ:

- учет – идентификация и инвентаризация существующих и вновь создаваемых документов и записей;
- определение политик – разработка, утверждение и обеспечение соблюдения политик ведения, оборота, хранения и уничтожения документов и записей;
- классификация документов и записей;
- хранение физических и электронных документов и записей (текущее и архивное);
- получение и распространение – регулирование доступа к документам и записям, их тиражирования и распространения в соответствии с установленными политиками и правилами, стандартами информационной безопасности и защиты данных, распоряжениями руководства организации и нормативно-правовыми требованиями;
- сохранение и уничтожение – своевременное архивирование и уничтожение документов и записей в соответствии с нуждами организации, законами, нормами и правилами.

Управление контентом

Управление контентом (content management) включает процессы, методы и технологии упорядочения, классификации и структурирования информационных ресурсов с целью обеспечения возможности их хранения, публикации и многократного многоцелевого использования.

Жизненный цикл контента может быть высокоактивным и предусматривать ежедневные изменения посредством контролируемых процессов создания, добавления или изменения информации. Существует также статичный контент, вовсе не меняющийся или изменяемый крайне редко и в минимальных пределах. В свою очередь, управление контентом может варьироваться от строго формализованного (в соответствии с жесткими правилами хранения, доступа, обращения и аудита, контролем сроков хранения и ликвидации) до полностью неформального добавления и изменения контента пользователями.

Если управление контентом ведется в масштабах организации, такой подход называется управлением корпоративным контентом (Enterprise Content Management, ECM).

Контролируемые словари

Контролируемым словарем (controlled vocabulary) называют определенный перечень слов, которые допустимо использовать в индексах, названиях категорий, документов, файлов и иных объектов, а также тегах метаданных с целью обеспечения возможности поиска, извлечения и просмотра контента. Подобный регламентированный лексикон необходим также и для систематизации документов, записей и контента в

478 В ГОСТ Р ИСО 15489-1-2019 (официальный перевод стандарта ISO 15489-1:2016 на русский язык) слово record переводится как «документ». Это соответствует распространенной практике перевода международных стандартов в области управления документами на русский язык*. Однако такой перевод может вносить некоторую путаницу**.

* Храпцовская Н. Стандарты ISO в области управления документами // Information Management, 02–2013: 10–22. – URL: https://eos.ru/upload/Information%20Management_13-02_10-22.pdf.

** Стрелкова Е. Как переводить слово record. Блог-журнал для партнеров и заказчиков EMC Documentum в России и СНГ, 2010. – URL: <http://my-documentum.blogspot.com/2010/11/record.html>.

Стрелкова Е. Управление документами и управление записями – найдем 6 отличий. Блог-журнал для партнеров и заказчиков EMC Documentum в России и СНГ, 2010. – URL: <https://my-documentum.blogspot.com/2010/11/6.html>.

каталогах библиотек. Сложность структуры словарей может варьироваться от простого списка или меню до более сложных кругов синонимов или нормативных словарей, еще более сложных таксономий и вплоть до сложнейших онтологий и тезаурусов.

Электронное раскрытие информации

Юрисдикции многих государств предусматривают действие специальных регламентов, дающих возможность организациям представлять суду доказательства в электронной форме. Электронные документы обычно маркированы метаданными (в отличие от многих бумажных, которые могут как иметь, так и не иметь данные об их происхождении), которые позволяют использовать их как важную часть доказательной базы. Процедура электронного раскрытия информации (e-discovery) позволяет выявлять электронные записи, которые могут быть предъявлены в судах различных инстанций и юрисдикций в качестве документальных доказательств.

14.2.4. Контекстная диаграмма области знаний и уровни зрелости функции «Управление документами и контентом»

Контекстная диаграмма области знаний «Управление документами и контентом» представлена на рисунке 14.7.

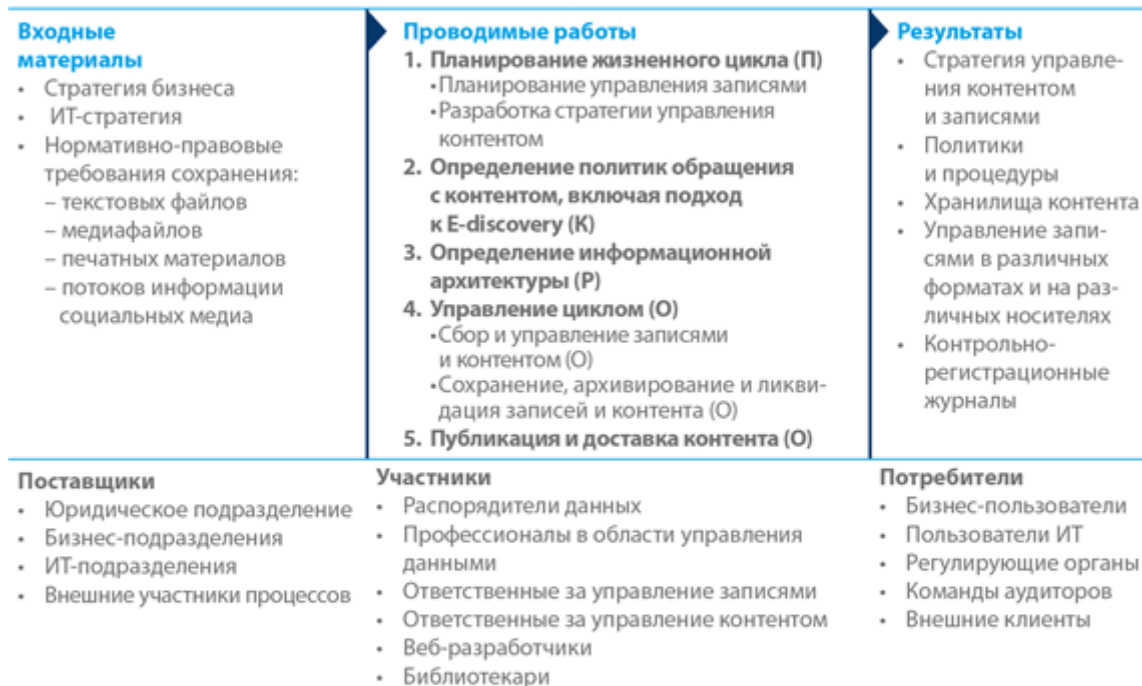
УПРАВЛЕНИЕ ДОКУМЕНТАМИ И КОНТЕНТОМ

Определение: планирование, реализация и контроль деятельности по управлению жизненным циклом неструктурированных (или полуструктурированных) данных и информации, представленных в любых форматах и на любых носителях

Цели:

1. Соблюдение требований действующего законодательства и соответствие ожиданиям клиентов в отношении управления записями
2. Обеспечение эффективного хранения, извлечения и использования документов и контента
3. Обеспечение интеграции структурированного и неструктурированного контента

БИЗНЕС-ДРАЙВЕРЫ



ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы • Маркировка с помощью метаданных • Использование формата XML • Карта данных • Сценарные планы • Инфографика	Инструменты • Офисное ПО • Системы управления корпоративным контентом (ECM) • Средства управления контролируемые словари / метаданными • Вики-ресурсы (управление знаниями) • Средства работы с визуальными медиа • Социальные медиа • Технологии e-discovery	Метрики • Метрики для аудита соответствия • Возврат инвестиций • Метрики использования • KPI управления записями • E-discovery KPI • Метрики для оценки ECM: – оценки программы – оценки операций

(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 14.7.

Контекстная диаграмма. Управление документами и контентом*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Для успешного управления документами, записями и другими видами совместно используемого контента требуется:

1. Четкое планирование, включая формирование политик для различных видов доступа и обработки.
2. Обеспечение возможности управления терминологией, включая онтологии и таксономии, необходимой для организации, а также хранения и извлечения различных форм контента.
3. Определение информационной архитектуры и метаданных, необходимых для поддержки контент-стратегии.
4. Внедрение технологий, позволяющих управлять жизненным циклом контента – от создания или сбора контента до управления версиями и обеспечения безопасности.
5. Для записей решающее значение имеет политика хранения и удаления. Записи должны храниться в течение требуемого периода времени и быть уничтожены, как только будут выполнены требования к сроку их хранения. Пока записи существуют, они должны быть доступны соответствующим людям и процессам и должны доставляться по соответствующим каналам.

Для достижения этих целей организациям требуются системы управления контентом (Content Management System, CMS), а также инструменты для создания и управления метаданными, которые поддерживают работу с контентом. Кроме того, им необходимо внедрить функцию руководства для контроля за политиками и процедурами, которые обеспечивают эффективное использование контента и предотвращают его неправильное применение. Такое руководство позволяет организации последовательно и должным образом реагировать на судебные разбирательства.

Система управления корпоративным контентом (ECM) может представлять собой как единое платформенное решение, включающее все основные компоненты, так и набор приложений с различной степенью интеграции в единую систему (от полностью интегрированных до полностью самостоятельных). Компоненты или приложения могут находиться как по месту работы, так и в облачной среде.

На рисунке 14.8 представлены обобщенные характеристики уровней зрелости функции «Управление документами и контентом».

14.2.5. Влияние на ценность данных

По мере развития технологий создания, хранения и использования данных объемы информации, сохраняемой в электронном виде (Electronically Stored Information, ESI), стремительно растут. От того, насколько предусмотрительно и активно ведется управление накапливающимися документами и контентом, зависят такие способности организации, влияющие на повышение ценности ее данных, как:

- способность оперативно предоставлять пользователям актуальные версии документов и контента;
- способность проведения новых видов анализа деятельности организации за счет связывания различных видов контента с уже имеющимися структурированными данными (например анализ проблем в отношении эксплуатации конкретного продукта на основе разбора связанных с ним писем от клиентов);
- способность адекватно и оперативно откликаться на запросы в отношении электронного раскрытия информации.

Практики в области DCM входят в состав ключевых компонентов стратегии организации по работе с данными. Они также рассматриваются как существенный фактор в рамках каждого проекта тактического уровня. Результаты обратной связи и оценка предыдущего опыта используются для выработки лучших практик. Проявляется активное стремление к дальнейшему совершенствованию. Клиенты (или граждане, поставщики и т. п.) могут участвовать в документообороте, например предоставляя отзывы о руководствах пользователя, содержанием веб-сайта и т. д. Осуществляется проактивное управление контентом с помощью использования данных для прогнозирования потребностей потребителей (как внутренних, так и внешних).

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Практики в области DCM измеряются и контролируются на основе общеорганизационных стандартов, при этом все отделы демонстрируют соответствие с помощью KPI и метрик. Осуществляется мониторинг рабочих процессов, контролирующих обмен документами и контентом по всей организации.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

С целью поддержки стратегической линии в области управления документами и контентом для всех бизнес-единиц определены общеорганизационные политики и стандарты. Подразделения выстраивают свои практики на основе этих стандартов. Осуществляется централизованное управление документами и контентом, включая возможность совместного использования и обмена документами по всей организации. Налажены координируемые рабочие процессы для создания контента и управления им. Для определенных категорий контента (например, для руководств по эксплуатации) осуществляется централизованное управление контролируемым словарем.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

В рамках конкретных подразделений или проектов успешно внедряются, документируются и повторно применяются практики DCM. В отдельных частях организации могут даже существовать локальные политики и стандарты, но нет никаких свидетельств того, что они координируются на более высоком уровне. Это приводит к применению несовместимых подходов или дублированию усилий на уровне организации в целом.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации отсутствуют какие-либо свидетельства наличия последовательной стратегии в области управления документами и контентом (DCM), поскольку существуют значительные различия в подходах отдельных подразделений и бизнес-функций.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 14.8.

Обобщенные характеристики уровней зрелости функции «Управление документами и контентом»*

* Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

14.3. Большие данные и наука о данных

За понятиями «большие данные» и «наука о данных» стоят значительные технологические изменения, благодаря которым появилась возможность генерировать, хранить и анализировать колоссальные объемы данных, и эти объемы продолжают неуклонно расти. Специалисты научились использовать такие данные для моделирования, прогнозирования и влияния на поведение людей, а также получения углубленных представлений о широком спектре важнейших предметов, включая статистику здравоохранения, управления природными ресурсами и экономического развития

Хотя значительная часть больших данных относится к категории неструктурированных и полуструктурированных, процессы в цепочках их поставок имеют существенно более сложную специфику, чем рассмотренные нами в предыдущем разделе особенности управления документами и контентом.

14.3.1. Определение функциональной области «Большие данные и наука о данных»

Чтобы охарактеризовать рассматриваемую область, приведем некоторые определения из ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии. Большие данные. Обзор и словарь».

Под большими данными (big data) понимаются большие массивы данных, отличающиеся главным образом такими характеристиками, как объем, разнообразие, скорость обработки и вариативность, которые требуют использования технологии масштабирования для эффективного хранения, обработки, управления и анализа (кроме того, термин «большие данные» широко применяется в различных значениях, например в качестве наименования технологии масштабирования, используемой для обработки больших массивов данных).

Таким образом, определение больших данных опирается на следующие ключевые понятия:

- **массив данных** (dataset) – идентифицируемая совокупность данных, к которой можно получить доступ или скачать в одном или нескольких форматах;
- **объем данных** (data volume) – количественная характеристика данных, влияющая на выбор ресурсов для вычислений и хранения, а также на управление данными в процессе обработки (объем данных становится важным при работе с большими массивами данных);
- **разнообразие данных** (data variety) – диапазон форматов, логических моделей, временных шкал и семантики массива данных (данное понятие отражает нерегулярность и разнородность структур данных, навигации по структурам, запросов и типов данных);
- **скорость обработки данных** (data velocity) – скорость потока, с которой данные создаются, передаются, сохраняются, анализируются или визуализируются;
- **вариативность данных** (data variability) – изменения в скорости передачи, формате или структуре, семантике или качестве массива данных.

Под наукой о данных (data science) понимается извлечение практических знаний из данных посредством исследования или создания и проверки гипотез.

Наука о данных изучает полный жизненный цикл аналитики данных. Аналитика данных (data analytics) – это составное понятие, охватывающее получение, сбор, проверку и обработку данных, включая их количественную оценку, визуализацию и интерпретацию.

Аналитика данных используется для представления объектов, описываемых данными, с целью прогнозирования конкретных ситуаций и формирования пошаговых рекомендаций при решении задач. Закономерности, полученные посредством аналитики, используются в различных целях, таких как принятие решений, проведение исследований, обеспечение устойчивого развития, проектирование, планирование и т. д.

В принципе, понятие «наука о данных» используется для обозначения хорошо известной дисциплины – прикладной статистики (applied statistics). Отличия обуславливаются тем, что вычислительные мощности, необходимые для выявления статистических закономерностей, сегодня выросли настолько, что способствовали появлению больших данных и реализации технологий их статистико-аналитической обработки.

До недавнего времени углубленный анализ колоссальных массивов данных был невозможен по технологическим причинам, и аналитикам приходилось полагаться на ограниченные по размерам статистические выборки или иные средства приблизительной оценки. С ростом вычислительных мощностей ученые научились накапливать и обрабатывать более объемные массивы данных и применять к ним комплексные методы анализа, позаимствованные из прикладной математики, статистики,

информатики, обработки и преобразования сигналов, теории вероятностей, распознавания образов, машинного обучения, моделирования неопределенности, визуализации данных и других прикладных областей знания с целью углубленного изучения и предсказания поведения систем на основе массивов больших данных. Иными словами, наука о данных нашла новые способы анализа данных и извлечения из них ценности.

Специалистов, которые исследуют данные, строят предиктивные (predictive) и предписывающие (prescriptive) модели, а также модели машинного обучения (machine learning), проводят на их основе анализ и осуществляют внедрение полученных результатов в интересах заинтересованных сторон, стали теперь называть «учеными в области данных» или «учеными по данным» (data scientists)⁴⁷⁹.

Важно понимать, что рассмотренные нами отличительные характеристики больших данных предъявляют новые требования к методам управления данными. Для использования преимуществ больших данных необходимо изменить привычные методические подходы. Большинство хранилищ данных используют традиционную реляционную модель. Большие данные, как правило, в виде такой модели не представлены. В большинстве хранилищ данных обработка тесно связана с процедурами ETL (извлечение, преобразование, загрузка). В решениях для обработки больших данных (в частности, в так называемых «озерах данных») используется концепция ELT, т. е. загрузка и последующее преобразование. Не менее важно и другое: скорость и потоки загрузки в случае сбора больших данных столь велики, что стандартные подходы к критически важным аспектам управления данными – интеграции, управлению метаданными, обеспечению качества данных – становятся неприемлемыми, и возникает необходимость в выработке и реализации принципиально новых решений еще и в этих областях⁴⁸⁰.

14.3.2 Цели и бизнес-драйверы

Организации осуществляют деятельность в области больших данных и науки о данных со следующими целями:

- раскрытие связей между данными и бизнесом;
- итеративное включение источников данных в среду организации;
- выявление и анализ новых факторов, которые могут оказывать влияние на бизнес;
- публикация и визуализация достоверных данных в подходящей и этичной форме.

В своей основе цели деятельности в области больших данных и науки о данных достаточно близки к целям деятельности в области ведения хранилищ данных и бизнес-аналитики (см. раздел 14.1.2). При этом имеется существенное отличие.

Традиционная бизнес-аналитика (BI) подобна «зеркалу заднего вида», поскольку описывает тенденции, выявленные по результатам изучения структурированных ретроспективных данных. Иногда выявленные закономерности бизнес-аналитики используются и для прогнозирования, но уверенности в надежности таких прогнозов нет, поскольку это всего лишь экстраполяции в будущее прошлых тенденций, которые в любой момент могут измениться.

С развитием технологий обработки больших данных и методов науки о данных организации приобретают способность смотреть вперед – «через лобовое стекло». Возможность прогнозирования на основе моделей, в том числе в режиме, близком к реальному времени, с использованием разнородных данных из множества различных источников помогает организациям лучше понимать направления своего развития.

Главный драйвер развития в организации работ в области сбора и исследования больших данных – стремление к обнаружению скрытых бизнес-возможностей посредством всесторонней аналитической

⁴⁷⁹ Можно сказать, что любой ученый по данным – это аналитик данных, но не любой аналитик данных – ученый по данным. В дополнение к навыкам аналитика, ученому по данным требуется углубленная математическая подготовка, понимание статистических методов, а также всего, что связано с их применением на практике*.

* Ханжина Ю. Л. Аналитик данных и data scientist: в чем отличие? Национальный исследовательский университет «Высшая школа экономики». Факультет компьютерных наук. Центр непрерывного образования, 2021. – URL: <https://cs.hse.ru/news/454940789.html>.

проработки массивов данных с использованием широкого спектра диверсифицированных алгоритмов. Большие данные стимулируют инновации, поскольку объемы и разнообразие массивов, доступных для исследования, безостановочно растут и все эти данные можно использовать для определения моделей прогнозирования нужд потребителей и создания персонализированных презентаций продуктов и услуг. Наука о данных способствует повышению производительности и результативности обработки больших данных. Алгоритмы машинного обучения помогают автоматизировать сложные по структуре и ресурсоемкие комплексы рабочих процессов, способствуя повышению эффективности работы организации, снижая затраты и минимизируя риски⁴⁸¹.

14.3.3. Дата-инжиниринг и экосистема больших данных

В разделе 14.1.8 мы рассматривали архитектуру комплексной рабочей среды для областей DW/BI и работы с большими данными (см. рис. 14.4). В процессе обработки входящих потоков больших данных сначала осуществляется их загрузка в специальное хранилище – озеро данных (data lake), а затем проводятся работы по интеграции и исследованию данных с построением моделей.

Поскольку сведения в озере данных могут быть необработанными (сырыми) и поступать из источников, не относящихся к операционным информационным системам организаций, они не подходят для рядового бизнес-пользователя; скорее, озера данных предоставляют материал для работы ученых по данным и различного рода экспертов, проводящих подробный анализ данных.

В связи с этим возникает необходимость в такой важной области деятельности, как дата-инжиниринг.

Дата-инжиниринг (data engineering) – это комплексная деятельность по обеспечению возможности использования необработанных данных. Без подготовительных работ им было бы невозможно разобраться в огромных объемах больших данных. За выполнение таких работ отвечает отдельная группа специалистов – инженеры данных (data engineers).

Инженеры данных – это инженеры-программисты (software engineers), которые, как правило, отвечают за построение конвейеров данных (data pipelines) для объединения информации из разных систем-источников. Они интегрируют, консолидируют и очищают данные и структурируют их для использования в аналитических приложениях.

Инженеры данных работают совместно с учеными по данным, повышая прозрачность данных и позволяя организациям принимать более надежные бизнес-решения.

Объем данных, с которыми работает инженер данных, зависит от организации и особенно от ее размера. Чем крупнее организация, тем сложнее архитектура аналитики и тем за большее количество данных он будет отвечать. Некоторые отрасли обрабатывают данные более интенсивно, в том числе здравоохранение, розничная торговля и финансовые услуги.

Основная цель инженера данных – сделать данные легко доступными и оптимизировать экосистему больших данных своей организации. Поэтому инженер данных должен иметь обширные знания в области современных технологий хранения и обработки данных, поскольку экосистема больших данных может включать самые разнообразные компоненты (рис. 14.9).

В первую очередь следует выделить распределенные файловые системы. Они работают на нескольких серверах сразу, способны хранить файлы, превышающие по объему размер диска отдельного компьютера, ориентированы на параллельную обработку файлов (одновременно на нескольких компьютерах) и легко масштабируются⁴⁸².

481 Там же.

482 Наиболее популярная распределенная файловая система – Hadoop File System.

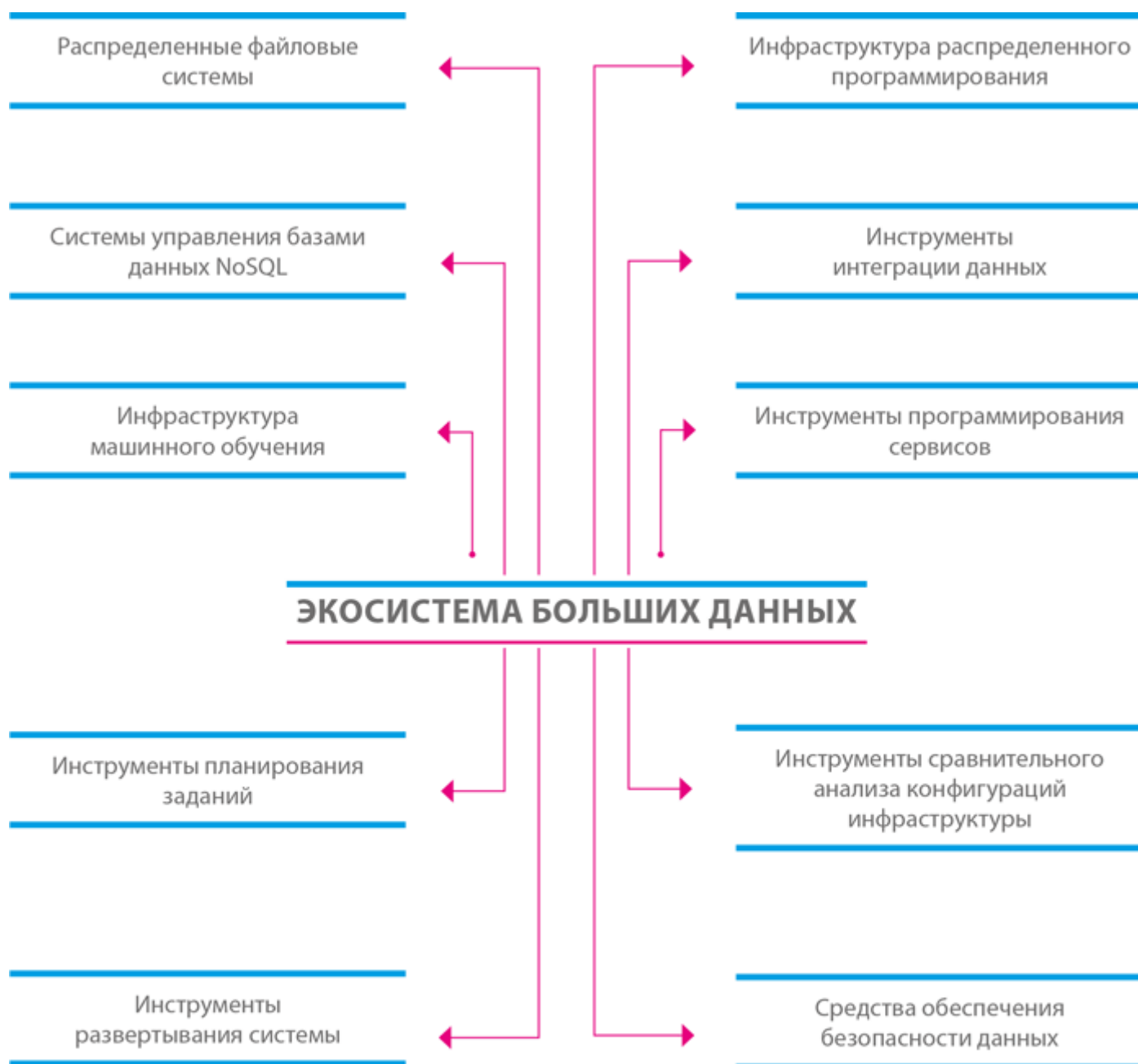


Рис. 14.9.

Экосистема больших данных*

* Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2018.

Для работы с данными в распределенной файловой системе должна быть использована специальная инфраструктура распределенного программирования⁴⁸³.

Хранение огромных объемов данных предполагает использование систем управления базами данных, специализирующихся на работе с такими данными и формировании запросов к ним. Традиционные реляционные СУБД, использующие язык запросов SQL (такие как Oracle или MySQL), плохо справляются

⁴⁸³ Распространенная модель программирования для распределенных параллельных вычислений на больших данных – MapReduce, предложенная компанией Google.

с большими объемами. Кроме того, в них отсутствуют средства обработки потоковых, неструктурированных и графовых (ориентированных на представление в виде графа) данных. Поэтому появились новые типы СУБД на основе нереляционных технологий, объединенные в категорию NoSQL (см. главы 11 и 12).

Данные в распределенной файловой системе перемещаются от источников к потребителям с помощью специальной инфраструктуры интеграции данных.

Когда данные доходят до потребителя, начинается их обработка с целью извлечения из них скрытой полезной информации и знаний. На этой стадии используются методы из области машинного обучения, статистики и прикладной математики. Необходимые для работы алгоритмы предоставляются инструментами, входящими в среду инфраструктуры машинного обучения.

С целью обеспечения всем заинтересованным системам (вне зависимости от их внутренней организации) унифицированный доступ к создаваемым приложениям, их реализуют в виде сервисов. Для этого используют специальные инструменты программирования и стандарты реализации (см. главу 12).

Для автоматизации повторяющихся операций и запуска заданий по событиям используются инструменты планирования заданий, созданные специально для работы с большими данными.

Инфраструктуру, обрабатывающую большие объемы данных, необходимо оптимизировать (это может принести существенную экономию). Оптимизация осуществляется с помощью инструментов сравнительного анализа конфигураций.

Развертывание новых приложений в кластерах больших данных можно облегчить с помощью инструментов, обеспечивающих автоматизацию установки и настройки.

Наконец, средства обеспечения безопасности, поддерживают функционирование приложения в рамках единой централизованной системы управления доступом.

14.3.4. Архитектурные компоненты аналитической среды организации и роли работающих с ними специалистов

Схема на рисунке 14.10 отражает архитектуру аналитической среды организации в более упрощенном виде, чем схема на рисунке 14.4. На ней выделено пять слоев.

Слой источников данных включает системы оперативной обработки транзакций (OLTP), поддерживающие операционную деятельность организации. Кроме того, в него могут входить различные приложения, подключаемые по API, а также датчики, внешние устройства и другие источники данных, подключаемые напрямую или с помощью сетевых протоколов.

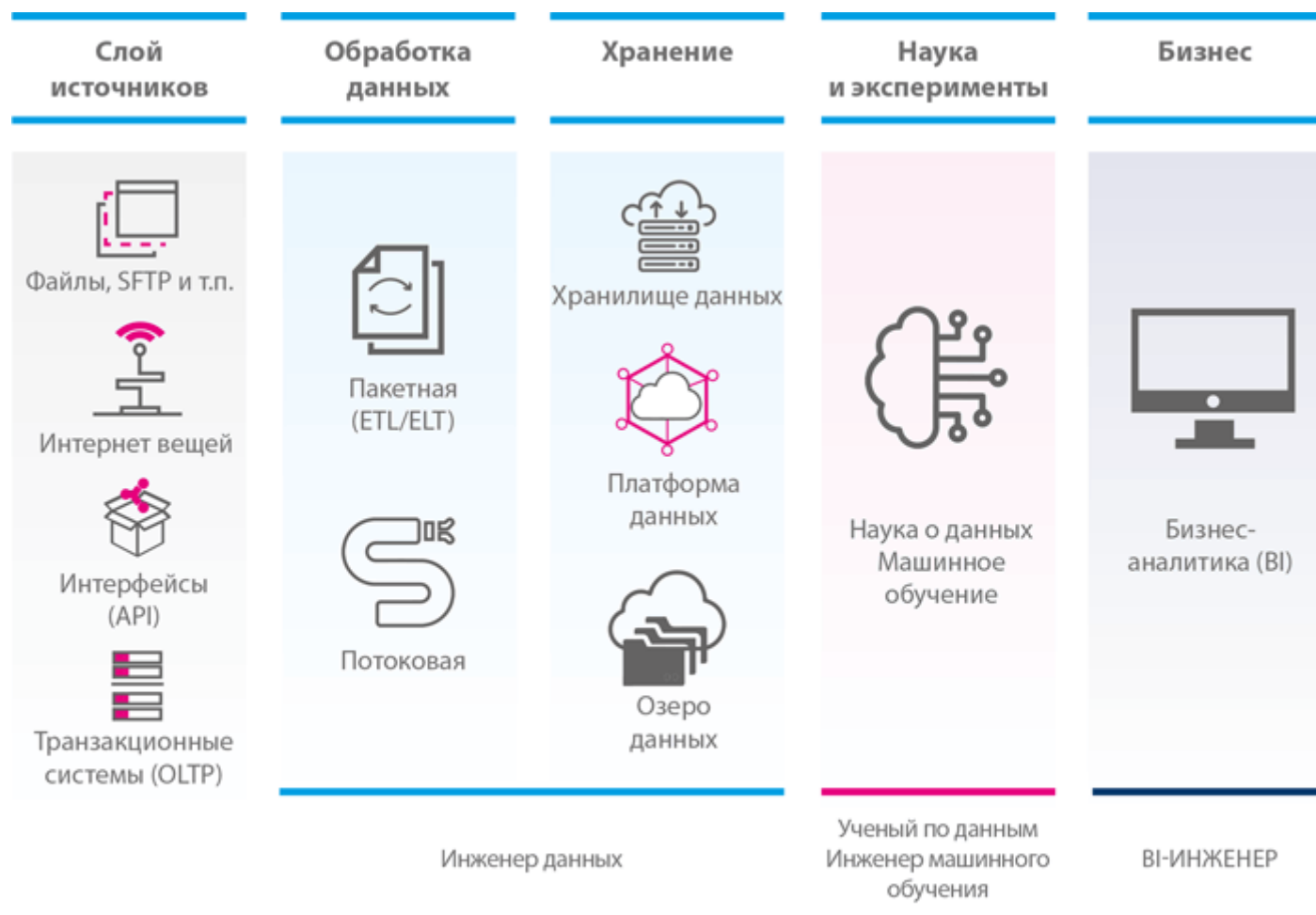


Рис. 14.10.

Упрощенное представление архитектуры аналитической среды организации*

* Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярчук. – М.: РАНХиГС, 2022.

Слой обработки данных выделен для обозначения операций, осуществляемых в пакетном режиме (с перерывами): ETL (извлечение – преобразование – загрузка) и ELT (извлечение – загрузка – преобразование), либо в потоковом (непрерывно).

Слой хранения может включать традиционное хранилище данных – Data Warehouse (DW), хранилище больших данных – озеро данных, либо современное хранилище, объединяющее DW и озеро данных, – платформу данных.

DW и озеро данных имеют схожую основную функцию (хранение данных для анализа), но различаются по своему назначению, структуре, видам хранящихся данных, а также их источникам и пользователям (см. табл. 14.2).

В DW собираются данные из бизнес-приложений для использования с конкретными целями. Перед хранением они должны быть очищены и упорядочены. При записи данные структурируют по предопределенной схеме (schema-on-write), что облегчает в дальнейшем доступ к ним.

Поскольку сведения, хранящиеся в DW, уже обработаны, их легче использовать для высокоуровневого анализа. Инструменты BI могут с ними легко оперировать, что упрощает использование хранилищ специалистами, не являющимися профессионалами в области работы с данными.

Таблица 14.2.

Сравнение хранилища данных и озера данных*

	Хранилище данных	Озеро данных
Виды данных	Структурированные обработанные данные из операционных баз данных, приложений и транзакционных систем	Структурированные, полуструктурированные и неструктурированные данные с вебсайтов, датчиков, от приложений и т. п.
Назначение	Предопределенное назначение: бизнес-аналитика, периодическая отчетность, визуализация данных	Может не иметь предопределенного назначения; обычно применяется в целях машинного обучения (machine learning), глубокого анализа (deep analysis), обнаружения информации (discovery)
Пользователи	BI-инженеры, бизнес-аналитики, аналитики данных	Инженеры данных, ученые по данным
Использование схемы	Схема при записи (schema-on-write)	Схема при чтении (schema-on-read)
Преимущества	Сгруппированные исторические данные, хранящиеся в одном месте с легким доступом для конечного пользователя	Данные, хранящиеся в своем изначальном формате, что обеспечивает для ученых по данным гибкость при анализе и разработке моделей на основе сведений из различных источников

* Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярчук. – М.: РАНХиГС, 2022.

Озеро данных – это обширное хранилище, в котором собираются необработанные данные в изначальном собственном формате. Одно из преимуществ озера данных – то, что оно может хранить данные различной структуры. Каждый сохраненный элемент данных помечен уникальным идентификатором и снабжен метаданными, чтобы при необходимости его можно было легко запросить. Данные в озере хранятся без предопределенной схемы – аналитики структурируют их только в момент чтения для конкретной задачи (schema-on-read). При построении озер данных целесообразно следовать существующим на сегодня передовым практикам⁴⁸⁴.

Сравнительная характеристика хранилища данных и озера данных представлена в таблице 14.2.

Для наполнения хранилища применяются процессы ETL или ELT, тогда как для озера данных – преимущественно ELT или потоковая обработка данных (стриминг).

484 Gorelik A. The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. – O'Reilly Media, Inc., 2019.

Таблица 14.3.

Основные роли специалистов, работающих с аналитической средой организации*

Роль	Функции
Архитектор данных (data architect)	Разрабатывает и участвует во внедрении корпоративных моделей данных (модели данных в масштабах организации), потоков данных и других компонентов архитектуры данных; также оценивает и дает рекомендации по закупкам технологий управления данными
Специалист по моделированию данных (data modeler)	Документирует определения и правила, связанные с бизнес-данными, и создает концептуальные, логические и физические модели данных с целью перевода их в конкретные реализации баз данных
Администратор баз данных	Управляет функционированием баз данных, обеспечивая выполнение требований соглашений об уровне обслуживания (service level agreements) в части доступности, производительности, затрат, а также безопасности и резервного копирования
Инженер по качеству данных (data quality engineer)	Проводит работы по обнаружению и очистке данных низкого качества; взаимодействует с разработчиками приложений и бизнес-пользователями с целью организации процессов, предотвращающих проблемы в области качества данных
BI-инженер (BI engineer)	Документирует потребности бизнеса в части анализа данных и отчетности; разрабатывает и поддерживает решения в области BI, отчеты и информационные панели (dashboards)
Инженер данных (data engineer)**	Создает конвейеры данных (data pipelines), платформы для обработки больших данных, решения по интеграции данных в базах данных, хранилищах данных и озерах данных (на основе как локальных (on-premises), так и облачных технологий)
Аналитик данных (data analyst)	Извлекает данные из бизнес-приложений и баз данных, анализирует их и создает отчеты и информационные панели с целью выделения тенденций и другой полезной информации для конечных пользователей
Ученый по данным (data scientist)	Собирает и анализирует данные из баз данных и других источников, используя алгоритмы машинного обучения и предиктивные модели; создает решения по визуализации данных для бизнес-пользователей
Инженер машинного обучения (ML engineer)	Ученый по данным, специализирующийся на машинном обучении

* Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярчук. – М.: РАНХиГС, 2022.

Если говорить о построении современной платформы данных, то в настоящее время известно несколько перспективных архитектурных концепций. В частности, выделяются подходы Modern Data

Architecture, Lambda Architecture и Data Mesh Architecture⁴⁸⁵.

Modern Data Architecture объединяет преимущества DW и озера данных. При этом следует заметить, что у Modern Data Architecture отсутствует четкий дизайн с точки зрения внедрения тех или иных решений. Концепция реализации во многом зависит от видения главного инженера проекта.

Lambda Architecture – решение, построенное в том числе на концепции озера данных, которое позволяет решать задачи, связанные с обработкой в режиме реального времени, обрабатывая данные за миллисекунды.

Data Mesh Architecture активно использует стриминг-технологии, объединяет пакетную и потоковую обработки данных, а хранит данные в облаке. Благодаря этому у организаций появляется возможность анализировать данные в режиме реального времени, снизив при этом затраты на управление инфраструктурой хранилища.

Два последних слоя на рисунке 14.10 выделены для обозначения деятельности в области науки о данных (ее осуществляют ученые по данным и инженеры машинного обучения) и деятельности в области BI (ей занимаются BI-инженеры).

В таблице 14.3 описаны основные роли специалистов, работающих с аналитической средой организации.

Деятельность в рамках слоев обработки и хранения данных обычно осуществляется инженером данных. Коротко рассмотрим ее на примере операций, выполняемых в ходе процесса ETL⁴⁸⁶.

Извлечение данных

На этом этапе данные извлекаются из одного или нескольких источников и подготавливаются к преобразованию. Отметим, что для корректного представления данных после их загрузки в хранилище из источников должны извлекаться не только сами данные, но и информация, описывающая их структуру, из которой будут сформированы метаданные для хранилища,

Преобразование данных

Чаще всего преобразование включает следующие шаги:

- *Преобразование структуры данных*

Данные из различных источников могут отличаться своей структурной организацией: соглашениями о назначении имен полей и таблиц, порядком их описания, форматами, типами и кодировкой данных. Перед передачей в хранилище их нужно свести к единой структуре.

- *Агрегирование данных*

Наибольший интерес для анализа представляют данные, обобщенные по некоторому интервалу времени, по группе клиентов или товаров. Такие обобщенные данные называются агрегированными (иногда агрегатами), а сам процесс их вычисления – агрегированием.

- *Перевод значений*

Часто данные в источниках хранятся с использованием специальных кодировок, которые позволяют сократить избыточность данных и тем самым уменьшить объем памяти, требуемой для их хранения. Так, наименования объектов, их свойств и признаков могут храниться в сокращенном виде. В этом случае перед загрузкой данных в хранилище требуется выполнить перевод сокращенных значений в более полные и понятные.

- *Создание новых данных*

В процессе загрузки в хранилище может потребоваться вычисление некоторых новых данных на основе существующих, что обычно сопровождается созданием новых полей.

- *Очистка данных*

Наличие «грязных» данных – одна из важнейших и трудно формализуемых проблем аналитических

⁴⁸⁵ Arenadata^a. Пять подходов к построению современной платформы данных. Блог компании Arenadata, 2021. – URL: <https://arenadata.tech/about/blog/5-ways-to-dwh/>.

⁴⁸⁶ Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям: Учебное пособие. 2-е изд., испр. – СПб: Питер, 2013.

технологий. Очистка данных – это процедура корректировки данных, которые в каком-либо смысле не удовлетворяют определенным критериям качества, т. е. содержат нарушения структуры данных, противоречия, пропуски, дубликаты или неправильные форматы.

Загрузка данных

Процесс загрузки заключается в переносе данных из промежуточных таблиц в структуры хранилища данных. От продуманности и оптимальности процесса загрузки данных во многом зависит время, требуемое для полного цикла обновления данных в хранилище, а также их полнота и корректность.

Следует заметить, что описанный здесь спектр операций, выполняемых на этапе преобразования данных, часто расширяется. Особенно при работе с большими объемами быстро поступающих данных, когда процесс ETL заменяется на ELT (сначала данные извлекаются и загружаются в конечную систему, и лишь после этого происходит их преобразование).

В частности, в ходе преобразования может возникнуть необходимость в группировке или разгруппировке данных (объединение или разъединение данных по какому-либо признаку), нормализации (преобразование диапазона изменений числового признака в другой, более удобный для применения в процессе анализа) и квантовании (разбиение диапазона возможных значений числового признака на заданное количество интервалов и присвоение попавшим в них значениям номеров интервалов или иных меток).

14.3.5. Контекстная диаграмма функциональной области «Большие данные и наука о данных» и уровни зрелости работы с большими данными

Контекстная диаграмма функциональной области «Большие данные и наука о данных» представлена на рисунке 14.11. Процесс осуществления деятельности в области науки о данных представляет собой последовательность итераций. Результаты предыдущей итерации служат исходными данными для следующей. Каждая итерация включает следующие работы⁴⁸⁷.

- **Определение стратегии и потребностей бизнеса в области изучения больших данных.** Формулировка требования к желаемым результатам с указанием измеримых материальных выгод от их выполнения.

- **Выбор источников данных.** Идентификация пробелов в имеющейся базе информационных ресурсов и поиск источников данных, которые позволят заполнить эти пробелы.

- **Получение и освоение источников данных.** Получение всех необходимых наборов данных или доступа к их источникам с целью загрузки.

- **Проработка гипотез и методов их проверки средствами науки о данных.** Исследование источников данных с помощью средств профилирования, визуализации, статистического анализа с целью уточнения требований. Определение алгоритма модели и необходимых типов входных и выходных данных или моделирование нескольких альтернативных гипотез и методов анализа (например, сравнительный анализ группировок данных, выявленных посредством кластеризации, и т. п.).

- **Интеграция и согласование данных для анализа.** Годность модели зависит еще и от качества источников данных. Следует использовать данные из надежных и достоверных источников. При необходимости, с целью повышения качества и полезности вводимых наборов, нужно применять средства интеграции, очистки и доработки данных.

- **Исследование данных с использованием моделей.** Использование средств статистического анализа и алгоритмов машинного обучения для выявления закономерностей на основе интегрированных данных. Регулярная проверка валидности модели и при необходимости внесение корректив в параметры модели и настройки алгоритмов самообучения. По мере накопления статистики – доработка самой модели. Машинное обучение подразумевает многократные прогоны через модель больших массивов реальных данных с целью проверки гипотез и внесения корректив в настройки алгоритмов (например, выявления выпадающих из общего статистического ряда значений). Также в процессе такой проработки окончательно

⁴⁸⁷ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

уточняются требования. Эволюция модели выверяется по изначально определенным метрикам пригодности или реалистичности результатов. С появлением новых гипотез могут потребоваться дополнительные наборы данных, а по результатам их проверки – новые модели, выходные данные и даже требования.

● **Внедрение и мониторинг.** Модели, которые выдают полезную информацию, можно переносить в производственную среду и использовать для текущего мониторинга ситуации с целью получения данных или, напротив, появления нежелательных тенденций, ставящих под угрозу эффективность текущей бизнес-модели. На этой стадии проекты по изучению данных превращаются в обычные рабочие проекты DW/BI и в среде DW обрастают всеми необходимыми техническими доработками и компонентами (процедурами ETL, правилами качества, основными данными).

На рисунке 14.12 приведены обобщенные характеристики уровней зрелости в соответствии с моделью зрелости использования цифровых технологий работы с большими данными в организации для достижения социальных и экономических эффектов (модель BD4DE-MM). Модель построена с учетом концептуальных положений методологии DECA для оценки развития цифровой экономики⁴⁸⁸.

⁴⁸⁸ В 2017 году Всемирным банком в сотрудничестве с Институтом развития информационного общества (ИРИО) и другими партнерами в России была разработана методика оценки уровня развития цифровой экономики в стране (Digital Economy Country Assessment, DECA) и апробирована на примере Российской Федерации.

БОЛЬШИЕ ДАННЫЕ И НАУКА О ДАННЫХ

Определение: сбор (больших данных), анализ и визуализация (наука о данных) множества разнородных данных различных видов с целью получения ответов на те вопросы, которые будут сформулированы лишь в процессе анализа

Цели:

1. Раскрытие связей между данными и бизнесом
2. Итеративное включение источников данных в среду организации
3. Выявление и анализ новых факторов, которые могут оказывать влияние на бизнес
4. Публикация (и визуализация) достоверных данных в подходящей и этичной форме

БИЗНЕС-ДРАЙВЕРЫ

Входные материалы	Проводимые работы	Результаты
- Стратегия и цели бизнеса - Дерево решений (собственное, приобретенное или позаимствованное) - ИТ-стандарты - Источники данных	1. Определение стратегии в области больших данных и потребностей бизнеса (П) 2. Выбор источников данных (П) 3. Получение доступа к источникам и загрузка данных (Р) 4. Разработка гипотез и методов (Р) 5. Интеграция/Согласование данных для анализа (Р) 6. Исследование данных (Р) 7. Внедрение и мониторинг (О)	- Стратегия и стандарты в области больших данных - План работ с источниками данных - Подключенные источники данных - Первичные гипотезы - Результаты анализа - План развития
Поставщики	Участники	Потребители
- Проектировщики архитектуры платформы для работы с большими данными - Специалисты в области науки о данных - Производители данных - Поставщики данных - Потребители информации	- Проектировщики архитектуры платформы для работы с большими данными - Эксперты в предметных областях - Специалисты в области науки о данных - Ведущие разработчики аналитических методов - Менеджеры в области управления данными - Специалисты по метаданным	- Бизнес-партнеры - Руководство бизнеса - Руководство ИТ

ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы	Инструменты	Метрики
- Объединение данных - Методы машинного обучения - Расширенное контролируемое обучение	- Решения на основе распределенных файлов - Колоночное сжатие - Архитектуры MPP без разделения ресурсов - Вычисления и базы данных в оперативной памяти - Алгоритмы в базе данных - Инструменты визуализации данных	- Показатели использования данных - Время ответа и производительность - Метрики загрузки и сканирования данных - Показатели эффективности и истории успешных внедрений

(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 14.11.

Контекстная диаграмма. Большие данные и наука о данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Модель BD4DE-MM предусматривает семь областей оценки зрелости, которые называются

размерностями (dimensions) или ключевыми факторами успеха (key success factors)⁴⁸⁹:

1. Стратегия и регулирование.
2. Кадры и лидерство.
3. Данные.
4. Инструменты и аналитика.
5. Инфраструктура и безопасность.
6. Организация работы.
7. Воздействие.

Организация выступает в роли лидера, масштабно использующего большие данные и внедряющего новые технологии работы с большими данными.

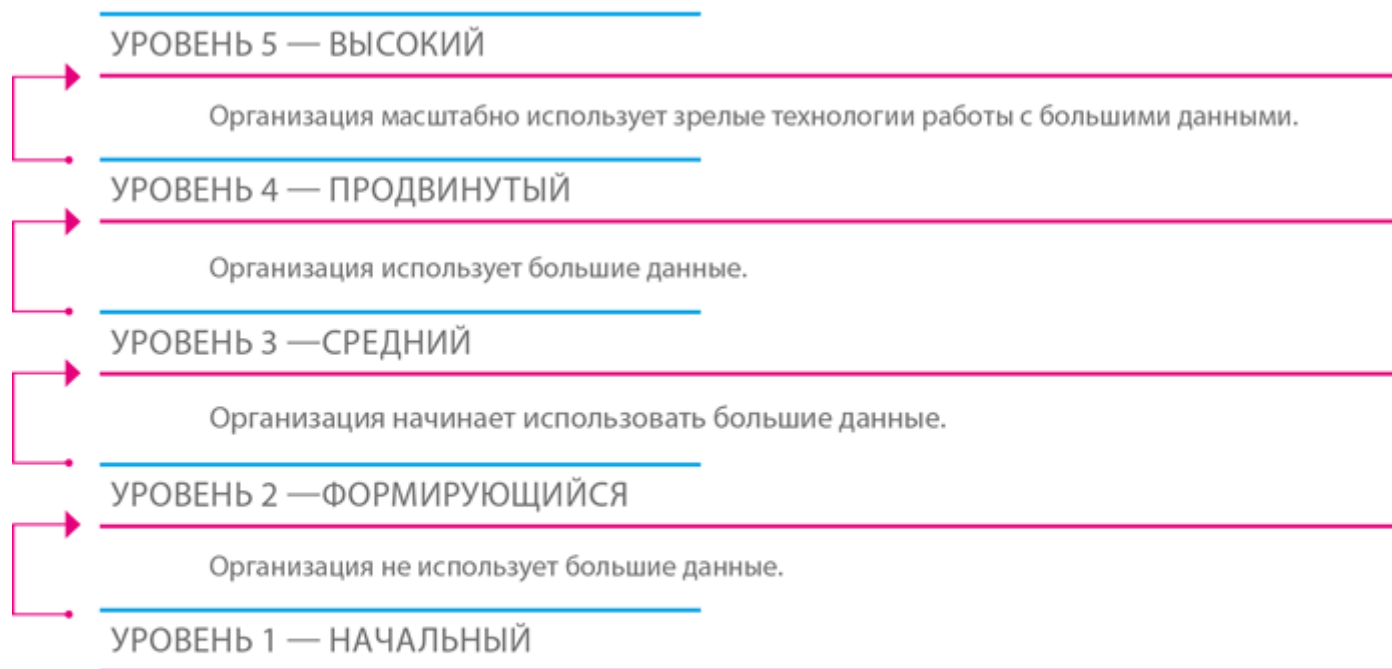


Рис. 14.12.

Обобщенные характеристики уровней зрелости использования цифровых технологий работы с большими данными (модель BD4DE-MM)*

* Ершов П. С., Катин А. В., Хохлов Ю. Е., Шапошник С. Б. Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, 4–5: 259–277. – URL: https://doi.org/10.52605/16059921_2021_04_259.

Для каждой из размерностей предусмотрен набор индикаторов оценки (assessment indicators). Например, для размерности «Стратегия и регулирование» в модель включены следующие индикаторы⁴⁹⁰:

1. Наличие в организации стратегии работы с большими данными.
2. Соответствие стратегии работы с большими данными положениям стратегии развития организации.
3. Наличие в организации плана реализации стратегии работы с большими данными.
4. Наличие в организации необходимых ресурсов (например, кадровых или финансовых) для реализации стратегии работы с большими данными.

⁴⁸⁹ Ершов П. С., Катин А. В., Хохлов Ю. Е., Шапошник С. Б. Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, 4–5: 259–277. – URL: https://doi.org/10.52605/16059921_2021_04_259.

⁴⁹⁰ Там же.

5. Наличие в стратегии работы с большими данными мероприятий, ориентированных на эксперименты с перспективными технологиями.
6. Соответствие деятельности организации требованиям нормативного правового регулирования работы с большими данными.
7. Соответствие деятельности организации международным стандартам работы с большими данными.
8. Соответствие деятельности организации принципам саморегулирования работы с большими данными.

14.3.6. Влияние на ценность данных

Согласно второму «закону» информации ее ценность возрастает с увеличением использования (см. главу 5). Таким образом, ценность данных не ограничивается одним конкретным случаем использования, их можно употребить многократно как с одной и той же целью, так и с разными. Применительно к большим данным особенно важен второй вариант.

В конечном счете ценность данных заключается в том, что можно получить от их всестороннего использования. Различные возможности использования служат альтернативами. Ценность данных определяется суммой таких вариантов – можно назвать это «альтернативной ценностью» данных. Раньше, после использования данных по основному назначению, было принято считать, что они свою миссию уже выполнили и их можно удалить. С появлением больших данных ситуация изменилась: данные обеспечивают отдачу еще долго после того, как их номинальная ценность уже извлечена. Можно выделить четыре эффективных способа раскрыть альтернативную ценность данных⁴⁹¹.

Простое повторное использование

Некоторые организации накапливают огромное количество данных, даже если не имеют в этом существенной необходимости или не практикуют их повторное использование. Так, например, операторы мобильной связи собирают информацию о местоположении своих абонентов, чтобы маршрутизировать их вызовы. Эти компании видят лишь узкое техническое назначение таких данных. Но их ценность значительно повышается при повторном использовании компаниями, которые распространяют персонализированную рекламу на основе местоположения.

Слияние наборов данных (искусственно созданные данные)

Согласно пятому «закону» информации ее ценность повышается при объединении с другой информацией (см. главу 5). Иногда скрытую ценность можно раскрыть, только объединив один набор данных с другим, возможно, совершенно непохожим. При анализе больших данных совокупность важнее отдельных частей, а при перекомпоновке совокупностей нескольких наборов данных получается еще более удачная совокупность.

Поиск массивов данных «2 в 1» (расширяемые данные)

Некоторые фирмы розничной продажи устанавливают в магазинах камеры наблюдения таким образом, чтобы не только обнаруживать злоумышленников, но и отслеживать передвижение клиентов по магазину и места, где они останавливаются, чтобы присмотреться. Такая информация полезна для разработки лучшей выкладки товаров в магазине, а также для оценки эффективности маркетинговых кампаний. Ранее камеры видеонаблюдения служили только для обеспечения безопасности и рассматривались не более чем статьей расходов. Теперь они рассматриваются как инвестиции, которые могут увеличить доход.

Дополнительные расходы на сбор нескольких потоков данных или намного большего числа точек данных в каждом потоке, как правило, невелики, поэтому имеет смысл собирать как можно больше данных, а также делать их расширяемыми, изначально рассматривая потенциальные варианты вторичного использования. Благодаря этому увеличивается альтернативная ценность информации. Суть в том, чтобы

⁴⁹¹ Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. – М.: Манн, Иванов и Фербер, 2014.

искать наборы «2 в 1», когда один и тот же массив данных, собранных определенным образом, можно применять в различных целях. Так эти сведения приобретают двойное назначение.

Учет амортизации ценности данных

Согласно третьему закону информации ее пригодность со временем снижается (см. главу 5). Информация с течением времени теряет часть своей первичной полезности. В таких условиях дальнейшее использование старых данных может не только не добавить ценности, но и фактически нивелировать пользу более новых данных. Разработка и применение моделей, которые позволяют выявить бесполезные сведения, чтобы своевременно их удалить, помогает повысить ценность имеющихся в распоряжении организации данных.

ПРАКТИЧЕСКИЙ ПРИМЕР

В рамках расширения задач по обеспечению доступности и обслуживанию данных «Телеком Дубль»:

- совершенствует архитектуру хранилищ данных;
- внедряет единую систему управления корпоративным контентом;
- развивает аналитическую среду компании, проводя работы по внедрению и развитию озера данных.

Большая работа проделана по оптимизации хранения профиля клиента. Та его часть, которая нужна онлайн (так называемый операционный профиль), перенесена на ИТ-инфраструктуру, которая держит высокие нагрузки по количеству выполняемых операций в секунду. Остальная часть остается в хранилище данных (аналитический профиль), где данные обновляются и пересчитываются в офлайне.

Важный аспект любого бизнеса – соответствие законам и нормативным актам, принятым в государстве, где юридическое лицо осуществляет свою деятельность. Документы компании описывают весь юридический процесс совершаемых операций – от транзакций по оплате пользования сервисами до работы с учредительными документами, – определяющих функционирование всех подразделений. Ведя централизованное управление документами и контентом, «Телеком Дубль» существенно повысила свою прозрачность.

Постепенно добавляя в описанный процесс все больше звеньев и детализируя его, мы можем проследить, каким образом каждый сотрудник влияет на достижение целей и задач компании. Это позволяет ему понять и визуально отследить собственный вклад в развитие бизнеса, а значит, дает возможность руководству ставить цели перед департаментами и отделами с любым уровнем детализации.

Литература к главе 14

- ГОСТ Р ИСО 15489-1-2019. Система стандартов по информации, библиотечному и издательскому делу. Информация и документация. Управление документами. Часть 1. Понятия и принципы.
- ГОСТ Р ИСО/МЭК 20546-2021. Информационные технологии. Большие данные. Обзор и словарь.
- Барсегян А. А., Куприянов М. С., Холод И. И., Тесс М. Д., Елизаров С. И. Анализ данных и процессов: Учеб. пособие. 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009.
- Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2018.
- Хохлов Ю. Е. Национальная политика работы с данными в Российской Федерации. Международные и национальные стандарты в сфере данных. Институт развития информационного общества, 2021. – URL: https://api.bigdata-msu.ru/media/uploads/2021/05/27/2021_05_25_hohlov.pdf.
- Bentley D. Business Intelligence and Analytics. Library Press, 2017.
- D-Russia.ru. Представлены 36 проектов национальных стандартов в области ИИ // D-Russia.ru, 2021. – URL: <https://d-russia.ru/predstavleny-36-proektov-nacionalnyhstandartov-v-oblasti-ii.html>
- Inmon W., Imhoff C., Sousa, R. Corporate Information Factory: 2nd Edition. Wiley Publishing, Inc., 2001.
- Loshin D. Business Intelligence: The Savvy Manager's Guide: 2nd Edition. Morgan Kaufmann, 2012.
- Loshin D. Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques,

NoSQL, and Graph: 1st Edition. Morgan Kaufmann, 2013.

• Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

• Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

• Управление данными в госсекторе. Навигатор для начинающих / под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярчук. – М.: РАНХиГС, 2022.

Глава 15. Базовая поддержка жизненного цикла данных

В рамочной структуре функций управления данными в привязке к их жизненному циклу, которая обсуждалась в главе 9 (рис. 9.4), выделено три направления деятельности по непосредственному управлению жизненным циклом данных:

- 1) планирование и проектирование данных;
- 2) обеспечение доступности и обслуживание данных;
- 3) практическое использование данных и расширение возможностей их применения для достижения целей организации.

В главах 11–14 мы рассмотрели области знаний (функции) по управлению данными, относящиеся к первым двум направлениям. Эти функции создают необходимые условия для осуществления деятельности на завершающей фазе жизненного цикла данных, включающей практическое использование данных и расширение возможностей их применения. Однако, прежде чем начать обсуждение этой фазы, следует остановиться еще на одной группе областей знаний, не менее важной для ее реализации. Как отмечалось в главе 9, в эту группу входят базовые функции, которые формируют основу для управления данными на протяжении всего их жизненного цикла. К ним относятся:

- 1) управление безопасностью данных;
- 2) управление метаданными;
- 3) управление качеством данных.

В главе 7, проводя сравнение элементов референтной модели управления цепями поставок (SCOR-модели) с цепочкой поставок данных, мы отметили, что перечисленные функции (вместе с функцией «Руководство данными») соотносятся с группой процессов «Предоставлять возможность». Поэтому, в отличие от глав 11, 12 и 14, в этой главе (как и в главе 10 «Руководство данными») нет специальных подразделов, посвященных анализу влияния обсуждаемых функций на ценность данных. Это влияние достаточно очевидно и состоит в первую очередь в максимальном содействии увеличению эффективности остальных функций, повышающих ценность данных на отдельных этапах их жизненного цикла.

15.1. Управление безопасностью данных

Специфика обеспечения безопасности данных (например, в отношении того, какие данные необходимо защищать) различается в разных отраслях и странах. Но цель соответствующих практик одна и та же: защитить информационные активы в соответствии с требованиями регулирующих органов и организаций, договорными обязательствами и бизнес-требованиями по безопасности и конфиденциальности.

15.1.1. Определение области знаний «Безопасность данных»

Область знаний «Безопасность данных» охватывает планирование, разработку и осуществление политик и процедур, обеспечивающих надлежащую аутентификацию, авторизацию и доступ пользователей, а также аудит данных и информационных активов⁴⁹².

⁴⁹² DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

В различного рода регламентирующих документах, связанных с безопасностью, вместо термина «безопасность данных» гораздо чаще используется термин «информационная безопасность». Для более полного описания рассматриваемой области приведем некоторые определения из ГОСТ Р ИСО/МЭК 27000-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология».

Под информационной безопасностью (ИБ) (information security) понимается сохранение конфиденциальности, целостности и доступности информации (этот термин может включать в себя и другие дополнительные свойства, такие как подлинность, подотчетность, неотказуемость и достоверность).

Отмеченные выше виды деятельности и свойства, характеризующие ИБ, определяются следующим образом:

- **аутентификация** (authentication) – обеспечение гарантии того, что заявленные характеристики субъекта и объекта являются подлинными;
 - **аудит** (audit) – систематический, независимый и задокументированный процесс, предназначенный для получения свидетельств аудита и объективной оценки аудиторами степени соблюдения критериев аудита;
 - **конфиденциальность** (confidentiality) – недоступность для неавторизованных лиц объектов или процессов;
 - **целостность** (integrity) – свойство сохранения правильности и полноты активов;
 - **доступность** (availability) – свойство, определяющее возможность использования объекта авторизованным субъектом по запросу;
 - **подлинность** (authenticity) – свойство, определяющее, что фактический субъект или объект совпадает с заявленным;
 - **неотказуемость** (non-repudiation) – способность удостоверять имевшее место событие или действие, которые в дальнейшем не могут быть поставлены под сомнение;
 - **достоверность** (reliability) – свойство соответствия предусмотренному поведению и результатам.
- В ГОСТ Р ИСО/МЭК 27000-2021 отмечается, что организации всех типов и размеров:
- собирают, обрабатывают, хранят и передают информацию;
 - осознают, что информация и связанные с ней процессы, системы, сети и персонал являются важными активами для достижения целей, стоящих перед организацией;
 - сталкиваются с рядом рисков, которые могут оказывать воздействие на функционирование активов организации;
 - принимают меры в отношении предполагаемого воздействия рисков, осуществляя внедрение мер обеспечения ИБ.

Вся информация, хранящаяся и обрабатываемая организацией, подвержена угрозам компьютерных атак, ошибкам, стихийным бедствиям (например, наводнению или пожару), а также это объект влияния уязвимостей, присущих ее использованию. Термин «информационная безопасность» связан с информацией, которую рассматривают как актив, представляющий собой ценность и требующий соответствующей защиты, например, от потери доступности, конфиденциальности и целостности. Обеспечение возможности санкционированного своевременного получения точной и полной информации способствует повышению эффективности бизнеса.

Защита информационных активов посредством определения, достижения, поддержания и улучшения ИБ необходима, чтобы обеспечить достижение намеченных организацией целей, а также поддерживать и повышать уровень соответствия законодательным нормам и репутацию организации. Эти скоординированные действия, направленные на внедрение соответствующих мер обеспечения информационной безопасности и обработку недопустимых рисков в области ИБ, широко известны как элементы менеджмента ИБ.

Так как риски ИБ и эффективность мер обеспечения ИБ меняются в зависимости от обстоятельств, организациям необходимо:

- контролировать и оценивать эффективность внедренных мер обеспечения ИБ и процедур;
- идентифицировать появляющиеся риски для их обработки;

- выбирать, внедрять и совершенствовать должным образом соответствующие меры обеспечения ИБ.

Чтобы установить взаимосвязи и скоординировать действия в рамках системы менеджмента информационной безопасности, каждая организация должна установить свою политику и цели для этой системы и эффективно достигать поставленных целей при ее функционировании.

15.1.2. Цели и бизнес-драйверы

Деятельность по управлению информационной безопасностью направлена на достижение следующих целей:

- обеспечение санкционированного доступа и исключение возможности несанкционированного доступа к информационным активам организации;
- обеспечение соблюдения нормативно-правовых требований и политик в отношении защиты информации о частной жизни, персональных и конфиденциальных данных;
- обеспечение соблюдения требований всех заинтересованных сторон в отношении защиты информации о частной жизни, персональных и конфиденциальных данных⁴⁹³.

Требования по обеспечению ИБ обусловлены следующими факторами (рис. 15.1).

1. Заинтересованные стороны. Организации должны выявлять и учитывать потребности в защите информации о частной жизни, персональных данных и конфиденциальных данных всех заинтересованных сторон, к которым могут относиться (в зависимости от типа и характера организации) клиенты, пациенты, студенты, граждане, поставщики, деловые партнеры и др. Все сотрудники организации несут ответственность за соблюдение требований, касающихся безопасности данных².

2. Нормативно-правовое регулирование. Речь идет о нормативно-правовом регулировании различных аспектов безопасности данных и интересов определенных заинтересованных сторон. Законы и правительственные постановления могут преследовать различные цели: одни ограничивают доступ к определенным данным, другие, напротив, призваны обеспечить открытость, прозрачность и подотчетность³.

3. Охрана интеллектуальной собственности и коммерческой тайны. В каждой организации имеются данные, которые можно расценивать в качестве предмета ее интеллектуальной собственности или коммерческой тайны. Такие данные нуждаются в защите. К примеру, клиентские базы данных помогают организации эффективно вести бизнес и получать преимущество перед конкурентами. В случае кражи, взлома системы хранения или уничтожения данных это преимущество будет сразу же утеряно⁴.

4. Потребности в санкционированном доступе к данным. Защита данных не должна ограничивать законный доступ к данным тех лиц, которые имеют на него право согласно действующему законодательству, равно как и санкционированный доступ к ним с целью использования, обслуживания, сопровождения, упорядочения и обработки в рамках бизнес-процессов⁵.

5. Договорные обязательства. Договорные обязательства и условия соглашений о неразглашении данных также сказываются на требованиях по обеспечению ИБ. Например, стандарт безопасности индустрии платежных карт требует от платежных систем, банков-эмитентов и коммерческих предприятий строго определенных мер по защите данных (например, обязательного шифрования паролей)⁴⁹⁴.

⁴⁹³ Там же.

⁴⁹⁴ Там же.

Безопасность данных

- О частной жизни и конфиденциальных данных клиентов
- О коммерческой тайне
- О деятельности партнеров
- О слиянии и поглощении

- Ограничения доступа к информации
- Требования открытости и подотчетности
- Права доступа для отдельных субъектов
- Другие требования



- Чрезмерная защита мешает вести бизнес
- Защита данных не должна мешать пользователям выполнять их работу
- Баланс простого и сложного

Безопасность данных

- О коммерческой тайне
- Об интеллектуальной собственности
- О потребностях клиентов
- Об отношениях с бизнес-партнерами и готовящихся сделках

Рис. 15.1.

Источники требований по обеспечению информационной безопасности*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Эффективные политики и процедуры в области безопасности данных гарантируют доступ к данным. При этом доступ предоставляется только тем, кому это положено в соответствии с их правами и обязанностями, а вот возможность несанкционированного доступа или изменения данных исключена.

Особую роль здесь играет тот факт, что обеспечение понимания и соблюдения потребностей и интересов всех сторон в отношении безопасности информации о частной жизни, персональных и конфиденциальных данных – важнейший аспект деятельности любой организации, поскольку ее отношения с клиентами, поставщиками и прочими заинтересованными сторонами строятся на доверии, ключевая составляющая которого – ответственное обращение с данными.

Основным драйвером работ по обеспечению ИБ выступает стремление минимизировать риски и обеспечить устойчивый рост бизнеса. Эффективное обеспечение безопасности данных снижает риски и дает дополнительные конкурентные преимущества. Безопасность данных можно рассматривать в качестве ценного актива организации.

Риски в области безопасности данных связаны с нормативно-правовыми требованиями, ответственностью руководства и/или владельцев перед компаниями, репутацией, исполнением юридических и моральных обязательств перед сотрудниками, партнерами и клиентами в части неразглашения их личных сведений, конфиденциальных данных и прочей нежелательной для раскрытия (чувствительной – sensitive) информации. С организаций могут взыскиваться крупные штрафы за несоблюдение установленных законом норм или неустойки за нарушение договорных обязательств. Утечки данных могут повлечь непоправимый репутационный ущерб и утрату доверия клиентов.

Рост бизнеса включает формулировку и достижение целей. Проблемные вопросы безопасности, утечки данных, равно как и необоснованные ограничения доступа к ним сотрудников, могут напрямую помешать успешному решению текущих задач.

Цели минимизации рисков и роста бизнеса могут быть согласованы и взаимно дополнять друг друга, если они объединены в комплексную стратегию управления информацией и обеспечения информационной

безопасности⁴⁹⁵.

15.1.3. Серия стандартов ИСО/МЭК 27XXX

Наиболее распространенным и общепризнанным в мире сборником рекомендаций в сфере обеспечения ИБ является комплекс международных стандартов серии ИСО/МЭК 27XXX, известный как семейство стандартов системы менеджмента информационной безопасности (СМИБ).

В состав семейства стандартов СМИБ входят взаимосвязанные стандарты, которые:

- определяют требования к СМИБ и к органам, сертифицирующим такие системы;
- обеспечивают непосредственную поддержку, содержат подробные рекомендации и интерпретацию общего процесса разработки, внедрения, поддержки и совершенствования СМИБ;
- содержат руководства по СМИБ для конкретных отраслей;
- содержат указания по оценке соответствия СМИБ.

Семейство стандартов СМИБ включает несколько ключевых структурных компонентов. К числу этих компонентов относятся прежде всего стандарты, определяющие:

- требования к СМИБ (ИСО/МЭК 27001);
- требования к органам по сертификации, осуществляющим сертификацию на соответствие ИСО/МЭК 27001 (ИСО/МЭК 27006);
- дополнительные требования, связанные с внедрением СМИБ в конкретных отраслях (ИСО/МЭК 27009).

Остальные стандарты предоставляют рекомендации по различным аспектам внедрения СМИБ, в том числе по общему процессу и управлению, а также специальные руководства для конкретных отраслей.

Базовый подход к управлению информационной безопасностью определяется двумя взаимосвязанными стандартами: ИСО/МЭК 27001⁴⁹⁶ и ИСО/МЭК 27002⁴⁹⁷ (рис. 15.2). Основную роль здесь играет стандарт ИСО/МЭК 27001, содержащий рекомендации по менеджменту ИБ в организации на основе широко используемого в корпоративной среде цикла управления качеством PDCA (Plan, Do, Check, Act). Стандарт ИСО/МЭК 27002 имеет скорее справочный характер, описывая набор возможных мер защиты информации, из которых организация может выбрать необходимые именно ей⁴⁹⁸.

Стандарт ИСО/МЭК 27001 дает рекомендации по функционированию СМИБ как комплексной системы, направленной на защиту информационных активов организации от угроз и, следовательно, минимизацию рисков. С точки зрения бизнеса СМИБ представляет собой одну из множества систем организации, к которой предъявляются определенные требования и которая должна оправдать ожидания и вернуть вложенные в нее средства.

В соответствии с рекомендациями стандарта, СМИБ включает в себя полный комплекс направлений деятельности по обеспечению информационной безопасности, в том числе организацию деятельности и управление рисками, а также непосредственное применение мер защиты информации. Выбирать те или иные способы защиты информации следует на основе оценки рисков ИБ: размера возможного ущерба от реализации угроз конфиденциальности, целостности и доступности информации. А также исходя из необходимости выполнения нормативных обязательств перед государством, партнерами и другими заинтересованными сторонами.

⁴⁹⁵ Там же.

⁴⁹⁶ Русскоязычная версия – ГОСТ Р ИСО/МЭК 27001-2021. Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования.

⁴⁹⁷ Русскоязычная версия – ГОСТ Р ИСО/МЭК 27002-2021. Информационные технологии. Методы и средства обеспечения безопасности. Свод норм и правил применения мер обеспечения информационной безопасности.

⁴⁹⁸ Прищеп С. Как избежать двойных стандартов в информационной безопасности. // ИКС (ИнформКурьер-Связь), 1. 2019: 84–88. – URL: <https://www.iksmedia.ru/articles/5558688-Kak-izbezhat-dvoynyx-standartov-v.html>.

**Рис. 15.2.**

Взаимосвязь стандартов ИСО/МЭК 27001 и ИСО/МЭК 27002

Таким образом, предлагаемый подход позволяет применять стандарт для реализации СМИБ в организациях любого масштаба и уровня нормативного регулирования.

Важно, что ИСО/МЭК 27001 совместим с другими стандартами систем менеджмента, такими как ИСО 9001, ИСО 14000, ИСО 31000, ИСО/МЭК 38500, ИСО/МЭК 20000, ИСО/МЭК 22301. Это позволяет использовать единый подход и принципы, общую терминологию, реализовать интегрированные процессы по направлениям контроля качества выпускаемой продукции, охраны окружающей среды, стратегического управления и управления ИТ-сервисами, обеспечения непрерывности деятельности организации, и, наконец, информационной безопасности. Что, в свою очередь, дает возможность построить структурированную и прозрачную систему менеджмента организации и повысить общую эффективность соответствующих процессов⁴⁹⁹.

15.1.4. Система менеджмента информационной безопасности

Система менеджмента информационной безопасности (СМИБ) включает в себя политику, процедуры, руководящие принципы и связанные с ними ресурсы, мероприятия, коллективно управляемые организацией в целях защиты ее информационных активов.

Эта система обеспечивает системный подход к созданию, внедрению, функционированию, мониторингу, анализу, поддержке и усилению ИБ организации для достижения бизнес-целей. Она основывается на оценке рисков и уровнях принятия рисков организацией, предназначенных для эффективной обработки рисков и управления ими. Анализ требований по защите информационных активов и применение соответствующих мер, обеспечивающих необходимую защиту этих активов, способствуют успешному внедрению СМИБ.

Для успешного внедрения СМИБ организации должны соблюдать следующие основные принципы⁵⁰⁰:

⁴⁹⁹ Там же.

⁵⁰⁰ ГОСТ Р ИСО/МЭК 27000:2021. Информационные технологии. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология.

- понимание необходимости использования СМИБ;
- назначение ответственности за обеспечение ИБ;
- обеспечение баланса между обязательствами руководства и потребностями заинтересованных сторон;
- повышение социальной значимости;
- оценивание рисков, чтобы применять необходимые меры обеспечения ИБ для достижения допустимых уровней рисков;
- обеспечение безопасности неотъемлемых элементов информационных сетей и систем;
- активное предупреждение и выявление инцидентов ИБ;
- применение комплексного подхода к менеджменту ИБ;
- регулярную переоценку уровня ИБ и внесение соответствующих изменений.

15.1.5. Разработка и сопровождение СМИБ

Организация должна предпринимать следующие шаги по разработке, мониторингу, поддержке и улучшению своей СМИБ:

- 1) определение информационных активов и связанных с ними требований ИБ;
- 1) оценка рисков ИБ и их обработка;
- 2) выбор и внедрение соответствующих мер обеспечения ИБ в отношении неприемлемых рисков;
- 3) мониторинг, поддержка и повышение эффективности мер обеспечения информационной безопасности, связанных с информационными активами организации.

Для гарантии эффективной непрерывной защиты информационных активов организации с помощью СМИБ необходимо постоянно повторять шаги 1–4, чтобы выявлять изменения в рисках, стратегии организации или бизнес-целях⁵⁰¹.

1. Определение информационных активов и связанных с ними требований ИБ

В рамках общей стратегии и бизнес-целей организации, ее размера и географического расположения требования ИБ можно сформулировать на основе анализа следующих факторов:

- идентифицированные информационные активы и их ценность;
- потребности бизнеса в обработке, обмене и хранении информации;
- юридические, нормативные и договорные требования.

Проведение систематической оценки рисков, связанных с информационными активами организации, включает в себя анализ угроз, уязвимостей, анализ потенциального воздействия любого инцидента ИБ и вероятности возникновения угрозы информационным активам, а также анализ потенциального воздействия любого инцидента ИБ на информационные активы. Расходы на соответствующие меры обеспечения ИБ будут пропорциональны предполагаемому влиянию риска на бизнес.

2. Оценка рисков ИБ и их обработка

Менеджмент риска ИБ требует должной оценки риска и метода его обработки. Это в свою очередь предполагает оценку затрат и преимуществ, законных требований, социальных, экономических и экологических аспектов, переменных, проблем заинтересованных сторон, приоритетов и других входных данных.

Оценка рисков должна выявлять, количественно оценивать и приоритизировать риски в сопоставлении с критериями принятия рисков и целями, представляющими важность для организации. Результаты оценки должны служить ориентиром и определять соответствующие управленческие меры и приоритеты для управления рисками ИБ и для внедрения мер обеспечения ИБ, выбранных для защиты от этих рисков.

Оценка риска должна включать в себя:

- систематический подход к оценке величины рисков (анализ рисков);

⁵⁰¹ Там же.

- процесс сравнения оцениваемых рисков с критериями риска для определения значимости рисков (оценка рисков).

Оценку рисков следует проводить регулярно. Это нужно, чтобы учитывать новые требования к ИБ, следить за ситуацией с рисками, например, в отношении активов, угроз, уязвимостей, воздействий, оценки рисков, а также в случае существенных изменений требований к ИБ. И, конечно, подобная оценка должна осуществляться по специальной методике, обеспечивающей сопоставимые и воспроизводимые результаты.

Чтобы оценка рисков ИБ была эффективной, она должна иметь четко определенную область применения и при необходимости проводиться совместно с оценкой рисков в других областях.

Стандарт ИСО/МЭК 27005 предоставляет рекомендации по менеджменту рисков ИБ, в том числе по оценке, обработке, принятию, мониторингу и проверке рисков, а также по связанным с рисками коммуникациям. Он также содержит примеры методик оценки рисков.

Прежде чем рассматривать вопрос, связанный с обработкой рисков, организация должна определить критерии, устанавливающие возможность принятия или непринятия риска. Риск может быть принят, если, например, определено, что он невысок, или не принят, если стоимость его обработки экономически нецелесообразна для организации. Такие решения следует документировать.

После оценки рисков необходимо принять решение об их обработке. Ниже перечислены возможные подходы к обработке рисков:

- применить соответствующие меры обеспечения ИБ, позволяющие снизить риски;
- сознательно и объективно принять риски при условии, что они четко соответствуют политике организации и критериям такого принятия;
- избегать рисков путем запрета действий, которые могут привести к возникновению рисков;
- распределить риски с другими сторонами, например, со страховщиками или поставщиками.

По тем рискам, в отношении которых было решено применить меры обеспечения ИБ, необходимо выбрать соответствующие меры и внедрить их.

3. Выбор и внедрение мер обеспечения ИБ

После определения требований ИБ, определения и оценки рисков ИБ для выявленных информационных активов и принятия решений по обработке рисков необходимо выбрать и внедрить меры обеспечения ИБ для снижения рисков.

Применяемые меры обеспечения ИБ должны способствовать снижению рисков до приемлемого уровня исходя из:

- требований и ограничений национального и международного законодательства и нормативных актов;
- целей организации;
- эксплуатационных требований и ограничений;
- стоимости их внедрения и эксплуатации с учетом снижения рисков при сохранении соразмерности требованиям и ограничениям организации;
- задач по мониторингу, оценке и повышению эффективности и действенности мер обеспечения ИБ в соответствии с целями организации;
- необходимости обеспечения баланса между инвестициями во внедрение и поддержку мер обеспечения ИБ и потерями, возможными в результате инцидентов ИБ.

Меры обеспечения ИБ, приведенные в стандарте ИСО/МЭК 27002, признаны передовой практикой, применимой к большинству организаций, и легко адаптируются к организациям различного размера и сложности. Другие стандарты из семейства стандартов СМИБ предоставляют рекомендации по выбору и применению мер обеспечения ИБ из ИСО/МЭК 27002 для СМИБ.

Меры обеспечения ИБ необходимо учитывать на этапе разработки требований к проектам и системам. В противном случае это может увеличить затраты и снизить эффективность решений, а в худшем случае – сделать невозможным достижение адекватного уровня безопасности. Меры обеспечения информационной безопасности могут быть выбраны из ИСО/МЭК 27002 или из других подходящих наборов мер обеспечения ИБ. Кроме того, для удовлетворения конкретных потребностей организации могут быть разработаны новые специальные меры обеспечения ИБ. Необходимо признать, что не все меры обеспечения информационной безопасности подходят для применения в информационных системах или

средах и практической реализации во всех организациях

В отдельных случаях для внедрения выбранного набора мер обеспечения ИБ требуется определенное время, и в течение этого времени уровень риска может быть выше допустимого в долгосрочной перспективе. Критерии риска должны охватывать допустимость рисков в краткосрочной перспективе, в период реализации мер обеспечения ИБ. Следует проинформировать заинтересованные стороны об уровнях риска, которые оцениваются и ожидаются в различные моменты времени, по мере постепенного внедрения данных мер обеспечения информационной безопасности.

Следует учитывать, что ни один набор мер обеспечения ИБ не может гарантировать полную ИБ. Необходимо внедрить дополнительные управленческие меры по мониторингу, оценке и повышению эффективности и действенности мер обеспечения ИБ в соответствии с целями организации.

Выбор и внедрение мер обеспечения информационной безопасности должны быть задокументированы.

4. Мониторинг, поддержка и повышение эффективности СМИБ

Организация должна поддерживать работоспособность и улучшать СМИБ посредством мониторинга и оценки эффективности в соответствии с политикой и целями организации, а также информировать руководство о полученных результатах для проверки. Цель такой проверки – удостовериться, что СМИБ включает в себя определенные меры обеспечения ИБ, применимые для обработки охватываемых ею рисков. Кроме того, на основе отчетов об этих областях мониторинга можно получить доказательства проверки и отслеживания корректирующих и предупреждающих мер, а также мер по улучшению ситуации.

Основная цель постоянного улучшения СМИБ – повышение вероятности достижения целей, связанных с сохранением конфиденциальности, доступности и целостности информации. Основная задача в этой сфере – поиск путей для совершенствования; не следует думать, что используемая практика управления ИБ достаточна или максимально эффективна.

Список мероприятий по улучшению весьма обширный:

- анализ и оценка существующей ситуации для выявления областей, нуждающихся в совершенствовании;
- постановка задач по совершенствованию;
- поиск возможных решений для достижения поставленных целей,
- оценка этих решений и выбор;
- внедрение выбранного решения;
- измерение, верификация, анализ и оценка результатов внедрения, чтобы определить, насколько достигнуты поставленные цели;
- официальное оформление изменений.

Полученные результаты перепроверяют по мере необходимости, чтобы наметить дальнейшие пути улучшения. Таким образом, улучшение – непрерывный процесс, действия повторяются с определенной частотой. Чтобы выявлять возможности улучшения, можно также использовать отзывы клиентов и других заинтересованных сторон, результаты аудитов и проверок СМИБ.

15.1.6. Контекстная диаграмма области знаний и уровни зрелости функции «Безопасность данных»

Контекстная диаграмма области знаний «Безопасность данных» представлена на рисунке 15.3.

Как и в случае с другими аспектами управления данными, лучше всего рассматривать информационную безопасность как корпоративную инициативу и делать это на протяжении всего жизненного цикла данных. Без скоординированных усилий бизнес-подразделения будут создавать различные индивидуальные решения для удовлетворения своих потребностей в области ИБ, что приведет к увеличению общих затрат при одновременном потенциальном снижении безопасности из-за несогласованной защиты. Неэффективная архитектура или процессы обеспечения безопасности могут дорого обойтись организациям из-за нарушений и потери производительности. Единая стратегия безопасности, которая согласована в масштабах всей организации и должным образом финансируется, снизит эти риски.

БЕЗОПАСНОСТЬ ДАННЫХ

Определение: планирование, разработка и осуществление политик и процедур, обеспечивающих надлежащую аутентификацию, авторизацию и доступ пользователей, а также аудит данных и информационных ресурсов организации

Цели:

1. Обеспечение санкционированного доступа и исключение возможности несанкционированного доступа к информационным активам организации
2. Обеспечение соблюдения нормативно-правовых требований и политик в отношении защиты информации о частной жизни, персональных и конфиденциальных данных
3. Обеспечение соблюдения требований всех заинтересованных сторон в отношении защиты информации о частной жизни, персональных и конфиденциальных данных заинтересованных сторон

БИЗНЕС-ДРАЙВЕРЫ

Входные материалы	Проводимые работы	Результаты
- Цели и стратегия бизнеса - Бизнес-правила и процессы - Нормативно-правовые требования - Корпоративная архитектура - Корпоративная модель данных	- Выявление требований по безопасности данных (П) - Определение политик безопасности данных (К) - Определение стандартов в области безопасности данных (Р) - Оценка текущих рисков безопасности данных (П) - Внедрение механизмов контроля и процедур (О)	- Архитектура безопасности данных - Политика безопасности данных - Стандарты безопасности данных о частной жизни и конфиденциальных данных - Механизмы контроля доступа к данным - Представления данных, соответствующие требованиям безопасности - Классификация данных с точки зрения безопасности - История операций аутентификации и доступа пользователей - Отчеты по результатам аудиторских проверок безопасности
Поставщики	Участники	Потребители
- Управляющий комитет по ИТ - Органы, отвечающие за корпоративную архитектуру - Регулирующие органы	- Распорядители данных - Команды по обеспечению безопасности - Внутренние аудиторы - Аналитики процессов	- Бизнес-пользователи - Аудиторы в области оценки нормативно-правового соответствия

ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы	Инструменты	Метрики
- CRUD-матрицы - Немедленное развертывание обновлений безопасности - Атрибуты безопасности в метаданных - Учет потребностей в безопасности данных в проектных требованиях - Санитизация документов	- Системы контроля доступа - Программные средства защиты - Технологии управления идентификацией - Программные средства обнаружения/предотвращения вторжений - Отслеживание метаданных - Маскировка/шифрование данных	- Эффективность внедрения мер безопасности - Осведомленность сотрудников в области ИБ - Защита данных - Инциденты безопасности - Предотвращение необоснованного распространения конфиденциальных данных

(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 15.3.

Контекстная диаграмма. Безопасность данных*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

На рисунке 15.4 отражено распределение деятельности в области управления ИБ по этапам

жизненного цикла данных.

На рисунке 15.5 представлены обобщенные характеристики уровней зрелости функции «Безопасность данных».



Рис. 15.4.

Распределение деятельности в области управления ИБ по этапам жизненного цикла данных*

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

В организации проводится регулярный аудит информационной безопасности с применением этичного хакинга и социальной инженерии. Результаты проверок используются в рамках постоянного совершенствования систем и процессов, а также при обучении сотрудников.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Эффективность практик обеспечения информационной безопасности оценивается с помощью ключевых показателей (KPI). Сведения об эффективности передаются команде по руководству данными. Политики и стандарты активно совершенствуются с целью поддержания соответствия текущим угрозам.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Для проведения стратегической линии в области безопасности данных в организации определены корпоративные политики и стандарты, распространяющиеся на все подразделения. По отношению к системам применяется согласованная политика информационной безопасности. Бизнес-процессы разработаны с учетом требований по защите данных.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

Имеются свидетельства применения документированных стандартов по безопасности данных, хотя нет гарантии, что они достаточно всеобъемлющи или, что возможно их согласованное внедрение по всем подразделениям. Встроенные в системы средства защиты принимаются по умолчанию в качестве действенных мер обеспечения безопасности без дополнительной оценки. Защита персональных данных осуществляется в первую очередь благодаря нерегламентированному обучению персонала и на основе здравого смысла.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

Практики обеспечения безопасности данных применяются в минимальном объеме и нерегулярно. Хотя существенных утечек данных, возможно, до сих пор не произошло, это скорее случайность, чем признанное правило. Необходимость защиты персональных данных не принимается во внимание.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

Рис. 15.5.

Обобщенные характеристики уровней зрелости функции «Безопасность данных»*

* Smith P., Edge J., Parry S., Wilkinson, D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

15.2. Управление метаданными

Внутри организации на разных уровнях и в различных подразделениях работает множество

сотрудников, у каждого из них собственный набор представлений о данных, которыми располагает организация, но никто не имеет и не может иметь исчерпывающего и достоверного представления о всей их совокупности. Поэтому требуется точный и подробный учет данных. Без ведения соответствующей документации организация рискует перестать понимать саму себя. Главным средством регистрации, формализации и упорядочения знаний о данных, имеющихся у организации, служат метаданные.

15.2.1. Определение области знаний «Метаданные»

Согласно ГОСТ Р ИСО/МЭК 11179-1-2010 «Информационная технология. Регистры метаданных. Часть 1. Основные положения» метаданные – это данные, которые определяют и описывают другие данные. Это означает, что данные, используемые подобным образом, становятся метаданными. Это происходит при конкретных обстоятельствах, для конкретных целей, с определенными перспективами, без которых данные не являются метаданными. Набор обстоятельств, целей или перспектив, для которых некоторые данные используются как метаданные, называют контекстом. Таким образом, метаданные являются данными о данных в некотором контексте⁵⁰².

В реальности к метаданным можно отнести очень широкий спектр сведений, включая информацию о технологических и бизнес-процессах, правила обработки данных, ограничения и определения логической и физической структуры данных.

Метаданные могут описывать не только данные как таковые (базы данных, элементы данных, модели данных), но и представляемые ими объекты (бизнес-процессы, системы и приложения, элементы ИТ-инфраструктуры), а также связи (отношения) между данными и объектами. Стоит отметить, что именно метаданные помогают организации правильно понимать смысл имеющихся в ее распоряжении данных, функционирование систем, структуру и содержание рабочих процессов. Они позволяют проводить оценку качества данных и неразрывно связаны с управлением базами данных и приложениями. Таким образом, метаданные необходимы для обеспечения возможности обработки, сопровождения, интеграции, хранения, защиты, проверки и контроля всех прочих данных организации.

Как и любые другие данные, метаданные нуждаются в управлении. С ростом способности организаций собирать и накапливать колоссальные массивы данных роль метаданных в сфере управления данными неуклонно возрастает. Чтобы быть «управляемой на основе данных» (data-driven), организация должна быть «управляемой на основе метаданных» (metadata-driven)⁵⁰³.

15.2.2. Цели и бизнес-драйверы

К основным целям управления метаданными относятся:

- управление задокументированными на уровне организации знаниями о данных в привязке к бизнес-терминологии с целью обеспечения единообразной трактовки данных всеми, кто их использует;
- сбор и интеграция метаданных из различных источников с целью обеспечения понимания пользователями сходств и различий между данными, поступающими из различных частей организации;
- обеспечение качества, согласованности, актуальности и защищенности метаданных;
- предоставление стандартных каналов доступа к метаданным всем потребителям данных (пользователям, системам, приложениям и процессам);
- выработка и утверждение собственных или контроль соблюдения предписываемых стандартов технических метаданных с целью обеспечения возможности обмена данными⁵⁰⁴.

В качестве главного бизнес-драйвера деятельности в области управления метаданными выступают

⁵⁰² Стандарты серии ГОСТ Р ИСО/МЭК 11179 определяют виды и качество метаданных, необходимых для описания данных, и специфицируют управление и администрирование этих метаданных в регистре метаданных (РМД). Эти стандарты применяют при формулировке представлений данных, понятий, смысла и взаимосвязей между ними для совместного использования людьми и машинами.

⁵⁰³ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁰⁴ Там же.

преимущества, которые обеспечивают для организации надежные и качественно управляемые метаданные. Здесь можно выделить:

- повышение доверия к данным за счет предоставления их контекста и поддержки возможности измерения качества данных;
- повышение ценности стратегической информации (в частности, основных данных) за счет ее многоцелевого использования;
- повышение эффективности работы информационных систем через выявление и устранение избыточных данных и процессов;
- своевременное выявление и отбраковку устаревших или неверных данных;
- оптимизацию планирования и проведения статистических исследований;
- лучшее взаимопонимание между потребителями данных и специалистами по ИТ;
- точность вводных данных, используемых для аналитического прогнозирования последствий, что способствует минимизации риска провала проектов;
- ускорение внедрения за счет сокращения времени, уходящего на разработку систем;
- снижение затрат на обучение и негативные последствия текучки кадров за счет исчерпывающей документации данных, включая контекст, источники и историю;
- выполнение требований действующего законодательства и надзорных органов.

Метаданные также способствуют согласованности и непротиворечивости данных и единообразному представлению информации, оптимизации потоков данных и рабочих процессов, надлежащей защите чувствительной информации, что особенно важно для отраслей с повышенными нормативно-правовыми требованиями.

При этом нужно понимать, что чем выше качество данных, тем выше их ценность для организации. В свою очередь, качество данных зависит от руководства данными. И вот тут на первый план выходят метаданные, без которых невозможно понимание данных в контексте функционирования организации. По сути, метаданные являются путеводителем по всем данным, имеющимся в распоряжении организации. Следовательно, управление метаданными должно быть безупречным. Плохо управляемые метаданные приводят к следующим негативным последствиям:

- появление избыточных данных и бессмысленных процессов управления ими;
- дублирующие друг друга, избыточные, устаревшие или вовсе не используемые словари, репозитории и иные хранилища метаданных;
- противоречивые определения объектов и элементов данных;
- неверные и противоречивые оценки рисков, соответствующих различным категориям данных, в том числе проистекающих от их нецелевого использования или утечки;
- конфликтующие между собой источники и версии метаданных и, как следствие, подрыв доверия пользователей к любым определениям данных, используемых в организации.

Хорошо организованное управление метаданными обеспечивает полное и согласованное представление об информационных ресурсах организации и способствует эффективному налаживанию взаимодействия между организациями при проведении совместной разработки приложений^{505, 506}.

15.2.3. Виды метаданных

Метаданные можно разделить на три основные категории:

- бизнес-метаданные;
- технические метаданные;
- операционные метаданные.

Бизнес-метаданные описывают содержание и состояние данных, а также детали, необходимые для реализации функций распоряжения данными.

Примеры бизнес-метаданных:

⁵⁰⁵ Там же.

⁵⁰⁶ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

- определения и описания наборов, таблиц и столбцов данных;
- бизнес-правила, правила преобразований, расчетные и логические формулы;
- модели данных;
- правила и результаты измерения показателей качества данных;
- расписания обновления данных;
- первоисточники и происхождение данных;
- стандарты данных;
- условные обозначения, используемые в системе записи и учета элементов данных;
- ограничения по допустимым значениям;
- контактная информация ответственных (например, владельцев или распорядителей данных);
- классы секретности и конфиденциальности данных;
- известные проблемы с данными;
- примечания по использованию данных.

Технические метаданные детально описывают всевозможные технические характеристики данных, систем их хранения и процессов перемещения данных между системами. Например:

- названия таблицы и столбцов таблицы, используемые в физической модели данных;
- свойства столбца;
- свойства объекта БД;
- права доступа;
- правила создания, замены, обновления и удаления записей (create, replace, update and delete; CRUD);
- физические модели данных, включая имена таблиц данных, ключи и индексы;
- задокументированные связи между моделями данных и физическими ресурсами;
- детализация операций по извлечению, передаче или загрузке данных (ETL);
- определения схем данных в файловых форматах;
- карты соотнесения данных между системами-источниками и адресатами;
- документация, описывающая происхождение данных, включая влияние изменений на информацию выше и ниже по потоку обработки;
- названия и описания используемых программ и приложений;
- расписания заданий по загрузке или обновлению контента и зависимостей между ними;
- правила резервного копирования и восстановления данных из резервных копий;
- права доступа, группы и роли пользователей.

Операционные метаданные детально описывают процессы обработки данных и управления доступом к ним. Например:

- журналы выполнения заданий пакетной обработки данных;
- история и результаты выгрузки выборок данных;
- сбои в расписаниях;
- результаты аудита, балансировки и контрольных измерений;
- журналы ошибок;
- структура, частота и время, скорость обработки запросов данных и отчетов;
- планы-графики исправлений, обновлений и выпуска новых версий и степень их соблюдения;
- правила резервного копирования, периодичности и сроков хранения резервных копий, порядок активации плана аварийного восстановления;
- требования и условия соглашений об уровнях обслуживания;
- схемы регистрации и распределения потоковой нагрузки;
- правила архивирования данных, сроки хранения архивов, правила обеспечения связности архивных данных;
- критерии окончательного удаления (утилизации) архивных данных;
- правила совместного доступа к данным;
- технические роли и обязанности, контактные данные.

Отметим, что на практике в процессе использования метаданных грани между описанными видами

оказываются довольно размытыми и не столь существенными^{507, 508}.

15.2.4. Архитектура метаданных

Как и другие данные, метаданные имеют свой жизненный цикл. Поэтому все решения по управлению метаданными включают следующие архитектурные уровни, соответствующие различным фазам жизненного цикла метаданных:

- создание или получение метаданных;
- хранение метаданных в одном или нескольких репозиториях;
- интеграция метаданных;
- доставка метаданных потребителям;
- использование метаданных;
- контроль и управление метаданными.

Для подключения к источникам, а также для сбора, хранения, интеграции и сопровождения метаданных и управления доступом к ним могут использоваться различные архитектурные подходы^{509, 510}.

● **Централизованный** : централизованная архитектура предусматривает единое хранилище метаданных, копируемых из различных источников. Организациям с ограниченными ИТ-ресурсами, как и стремящимся к максимально возможной автоматизации управления метаданными, такой вариант архитектуры, как правило, не подходит. В то же время организации, стремящиеся к согласованности метаданных, извлекают максимальную пользу от хранения их в централизованном хранилище.

● **Распределенный** : полностью распределенная архитектура предусматривает единую точку доступа к метаданным через портал, обеспечивающий извлечение запрашиваемых данных из систем-источников в режиме, близком к реальному времени. Центральное хранилище при такой архитектуре отсутствует. Вместо него в среде портала управления метаданными ведутся каталоги данных, содержащихся в системах-источниках, и действуют общие правила оптимизации обработки запросов. Обращение непосредственно к системам-источникам осуществляется посредством протоколов, используемых промежуточным ПО.

● **Гибридный** : гибридная архитектура сочетает в себе элементы, свойства и характеристики как централизованной, так и распределенной архитектур. Метаданные все так же поступают в центральный репозиторий непосредственно из систем-источников, но сохраняются они там выборочно. Обычно система управления таким хранилищем предусматривает сохранение критически важных стандартизованных элементов метаданных из систем-источников и последующее добавление дополнительных элементов по запросу пользователей, в том числе в ручном режиме из сторонних источников.

15.2.5. Контекстная диаграмма области знаний и уровни зрелости функции «Метаданные»

Контекстная диаграмма области знаний «Метаданные» представлена на рисунке 15.6.

На рисунке 15.7 отражено распределение деятельности в области управления метаданными по этапам их жизненного цикла.

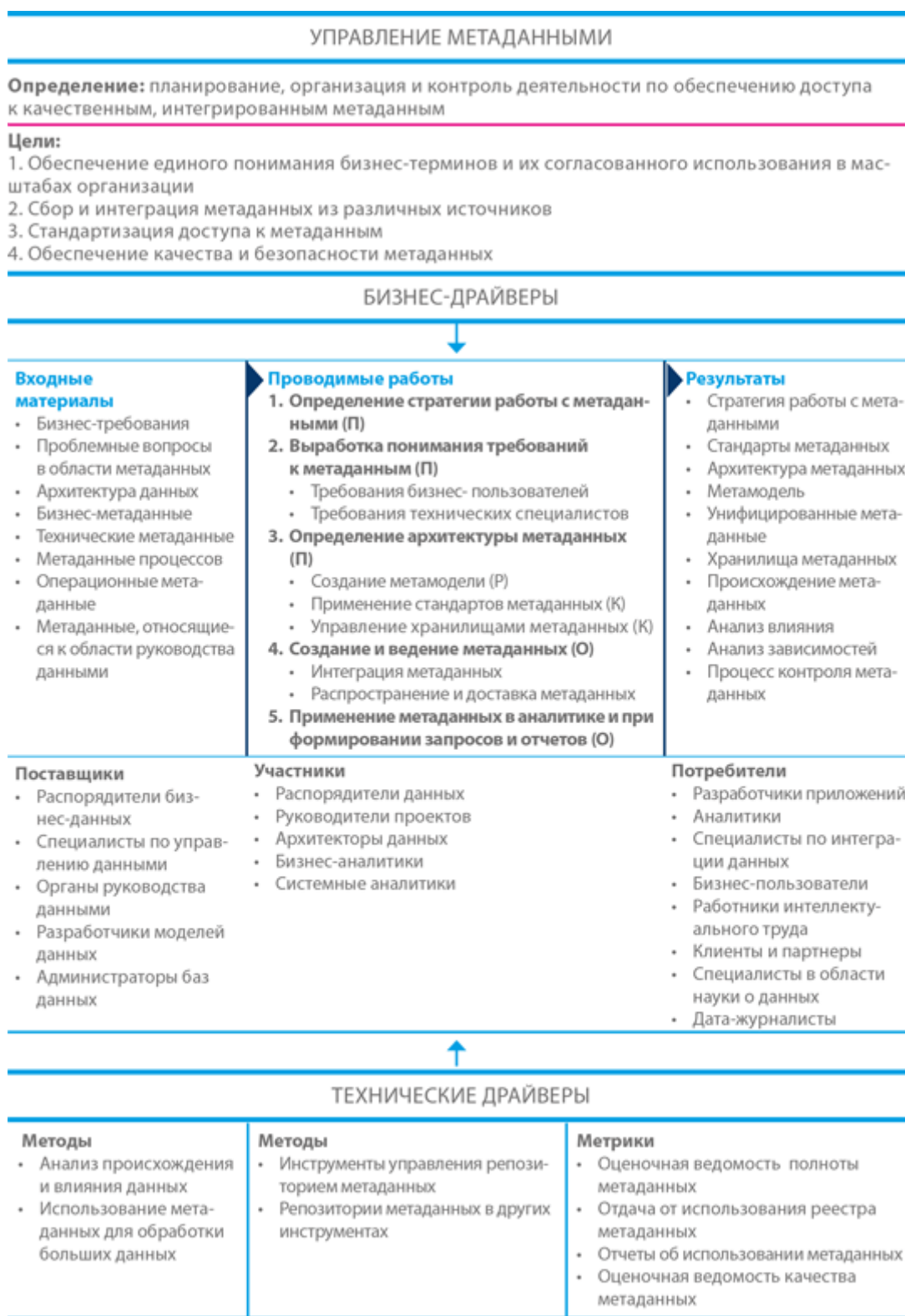
На рисунке 15.8. представлены обобщенные характеристики уровней зрелости функции «Метаданные».

⁵⁰⁷ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁰⁸ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

⁵⁰⁹ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵¹⁰ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.



(П) Планирование, (К) Контроль, (Р) Разработка, (О) Операции

Рис. 15.6.

Контекстная диаграмма. Метаданные*

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)



Рис. 15.8
Обобщенные характеристики уровней зрелости функции «Метаданные»

15.3. Управление качеством данных

Одна из трудностей управления качеством данных заключается в том, что ожидания в отношении качества данных не всегда известны. Бывает, что потребители просто неспособны их сформулировать. А порой случается и так: люди, отвечающие за управление данными, не отдают себе отчета в том, что к этим

данным могут быть применимы какие-то специфические требования.

Поскольку ни одна организация не может похвастаться безупречностью технологических и бизнес-процессов, а также практик управления данными, проблемы с качеством данных неизбежны. Однако в организациях, где реализована формальная система управления качеством данных, проблемы возникают реже и решаются проще, чем в организациях, где качество данных – дело случая.

15.3.1. Определение области знаний «Качество данных»

Термин качество данных (Data Quality, DQ) распространяется как на характеристики, связанные с высоким качеством данных, так и на процессы измерения или повышения качества данных.

Следует разделять эти два варианта использования термина и пояснять, что понимается под данными высокого качества.

Данные можно считать высококачественными в той мере, в которой они соответствуют потребностям и ожиданиям потребителей. То есть данные обладают высоким или низким качеством, если они, соответственно, пригодны или непригодны к использованию по назначению. Следовательно, качество данных зависит от контекста и потребностей потребителей данных.

Формальное управление качеством данных выстраивается по аналогии с непрерывным управлением качеством других продуктов. Качество данных контролируется на всех фазах их жизненного цикла посредством определения стандартов и встраивания механизмов обеспечения и контроля их соблюдения в процессы создания, преобразования и хранения данных, включая наборы измеримых показателей соответствия данных стандартам качества на всех этапах. Для внедрения комплексного подхода к обеспечению качества данных обычно требуется команда по реализации программы качества данных (data quality program team). Она отвечает за привлечение к участию и координацию действий профессионалов в области управления данными со стороны бизнеса и технических подразделений при проведении работ, обеспечивающих последовательное применение методов, которые гарантировали бы пригодность любых данных для использования по назначению. Более того, команде программы качества данных, вероятно, потребуются принять участие в серии проектов, прежде чем она сможет внедрить в организации устойчивые процессы с использованием передовых практик непрерывного управления качеством данных. Параллельно должны приниматься экстренные меры по устранению неотложных проблем.

Поскольку управление качеством данных предполагает управление их жизненным циклом, программа качества данных неизбежно накладывает определенные требования и ограничения на использование данных и предусматривает ответственность за обеспечение их соблюдения при осуществлении операционной деятельности. В обязанности участников команды программы качества данных может входить, например, составление отчетности об уровнях качества данных; участие в анализе данных и сборе статистики; выявление и приоритизация проблем с данными.

Помимо этого, команда программы качества данных отвечает за взаимодействие с потребителями данных при решении вопросов, касающихся обеспечения их потребностей, а с теми, кто задействован в создании, обновлении или удалении данных, – вопросов обеспечения соблюдения правил обращения с данными. Качество данных зависит от всех, кто с ними работает, а не только от профессионалов в области управления данными.

Так же как руководство и управление данными в целом, управление качеством данных осуществляется именно как систематическая программа, а не разовый проект. При этом программа качества данных включает и работы, проводящиеся на проектной основе, и плановую деятельность по сопровождению информационных систем и ресурсов, а также обеспечение эффективных коммуникаций и обучение⁵¹¹.

15.3.2. Цели и бизнес-драйверы

Программы качества данных преследуют следующие цели:

- выработка управляемого подхода к обеспечению соответствия данных нуждам их потребителей;
- определение стандартов и спецификаций механизмов контроля качества данных как составной

⁵¹¹ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

части жизненного цикла данных;

- определение и внедрение процессов измерения, мониторинга и учета уровня качества данных;
- выявление и поддержка использования возможностей по повышению качества данных посредством внесения изменений в системы и процессы, а также осуществление деятельности по проведению измеримых улучшений качества данных на основе требований их потребителей⁵¹².

Бизнес-драйверы, обуславливающие необходимость наличия формализованной программы качества данных:

- повышение ценности данных и информационных ресурсов организации и реальной отдачи от их использования;
- снижение рисков и издержек, обусловленных низким качеством данных;
- повышение эффективности и производительности в масштабах организации;
- защиту и укрепление репутации организации.

Организации, стремящиеся получать полноценную отдачу от имеющихся данных, прекрасно понимают, что высококачественные данные ценнее данных низкого качества. К тому же некачественные данные чреваты ущербом репутации, штрафами, упущенной прибылью, оттоком клиентов и негативными отзывами в СМИ. Обеспечивать высокое качество данных нередко предписывают также нормативно-правовые документы и отраслевые регламенты. Ну и наконец, некачественные данные влекут за собой и всевозможные прямые убытки. Приведем лишь некоторые примеры негативных последствий:

- ошибки в выставленных счетах;
- увеличение числа обращений в службу поддержки клиентов при одновременном снижении способности разрешать возникшие проблемы;
- упущенные возможности и, как следствие, падение оборота и выручки;
- задержка интеграции в процессе слияний и поглощений;
- повышенная уязвимость перед угрозой мошенничеств и злоупотреблений;
- убытки вследствие ошибочных бизнес-решений, сделанных на основе неверных данных;
- потеря бизнеса и/или клиентуры из-за неспособности подтвердить свою репутацию и/или кредитоспособность.

Однако высокое качество данных – не самоцель, а средство обеспечения организационного успеха. Достоверные данные не только снижают риски и издержки, но и повышают эффективность. Работая с надежными данными, сотрудники более оперативно и согласованно находят ответы на текущие вопросы и тратят меньше времени на поиск нужной информации и оценку ее пригодности, что оставляет им больше времени на глубокое осмысление данных с целью взвешенного принятия решений и качественного обслуживания клиентов⁵¹³.

15.3.3. Измерения качества данных

Измерениями качества данных (data quality dimension) называют измеримые свойства или характеристики данных, находящиеся в прямой связи с их качеством. Термин «измерение» сразу же приводит к ассоциативной аналогии с мерами свойств физических тел (таких как длина, ширина, высота). Измерения качества данных служат также источником терминологии, используемой для определения требований к качеству данных. Их же можно использовать для описания результатов как первичной оценки, так и текущих измерений качества данных. Для оценки качества данных организации нужно определить такие измерения, которые одновременно важны и для бизнес-процессов (и потому заслуживают рассмотрения) и поддаются объективной оценке. Измерения также служат базисной системой координат при определении правил оценки, которые, в свою очередь, напрямую соотносятся с потенциальными рисками, присущими критически важным процессам.

Часто выделяют следующие шесть ключевых измерений качества данных.

1. **Полнота** : отношение фактически имеющегося в хранилище объема данных к потенциально доступному (0–100 %).

512 Там же.

513 Там же.

2. **Уникальность** : ни одному реально существующему экземпляру предмета описания (объекта) не должно соответствовать более одной записи в рамках идентификации описываемых предметов/объектов.
3. **Актуальность** : степень отражения данными реального положения вещей на текущий момент.
4. **Годность** : определяется синтаксическим соответствием данных определениям (по формату, типу или диапазонам значений).
5. **Соответствие** : Степень соответствия данных реальным объектам или событиям, которые ими описываются.
6. **Согласованность** : отсутствие противоречий между различными представлениями одного и того же (согласно определениям) предмета или сущности.

Иногда рассматриваются и другие важные характеристики, влияющие на качество данных.

- **Полезность** : насколько понятны, доходчивы, релевантно определены, доступны и точны данные?
- **Своевременность реагирования** (в дополнение к актуальности): поддерживается ли возможность оперативного изменения данных без потери стабильности?
- **Гибкость** : насколько данные совместимы и сопоставимы с другими данными? Допускают ли группировку, классификацию и перепрофилирование? Достаточно ли просты в обращении?
- **Надежность** : организованы ли процессы руководства данными и обеспечения безопасности данных? Какова репутация данных, чем или как она подтверждается или удостоверяется?
- **Ценность** : имеется ли экономическое обоснование с анализом рентабельности или окупаемости затрат на управление данными? Оптимально ли используются данные? Все ли в порядке с защитой персональных, личных и конфиденциальных данных? Не допускается ли предприятием каких-то неправомерных действий или нарушений? Соответствует ли его деятельность корпоративному имиджу?

Единой универсальной классификации измерений качества данных до сих пор не выработано, однако вышеописанные формулировки содержат общие идеи. Измерения включают часть характеристик, оцениваемых по вполне объективно измеримым показателям (например, полнота, действительность, соответствие формату), и часть, которая в значительной степени зависит от контекста или субъективной интерпретации (полезность, надежность источника, репутация). Какие бы названия измерений ни использовались, основными аспектами качества данных являются: полнота (отсутствие недостающих данных); правильность (корректность, точность, достоверность); непротиворечивость (согласованность, целостность, уникальность), актуальность (своевременность обновления или реагирования); доступность; возможность использования (годность); безопасность (защищенность)^{514, 515, 516, 517}.

15.3.4. Жизненный цикл повышения качества данных

Большинство методологических подходов к повышению качества данных позаимствованы из теории управления качеством на производстве. В рамках такой парадигмы любые данные считаются конечным продуктом комплекса технологических процессов по переработке информационного сырья. Процесс создания данных может быть простым и одношаговым (сбор или получение), а может быть многоэтапным и включать целый ряд последовательных информационно-технологических операций: сбор данных, включение и накопление в хранилище, обобщение в витрине данных. На каждом этапе данные и их качество подвергаются риску: при сборе возможны ошибки; при передаче из системы в систему – потери, дублирования или искажения; при интеграции и накоплении, анализе или обобщении – методологические

⁵¹⁴ Там же.

⁵¹⁵ McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM). Morgan Kaufmann, 2008.

⁵¹⁶ Myers, D. The Value of Using the Dimensions of Data Quality, Information Management, August 2013, <http://bit.ly/2tsMYiA>.

⁵¹⁷ Sebastian-Coleman L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework (The Morgan Kaufmann Series on Business Intelligence), First Edition. Morgan Kaufmann, 2013.

ошибки и технические проблемы. Для повышения качества данных необходимо располагать возможностью оценки соответствия выходных данных ожиданиям, которые определяются, с одной стороны, фактическим содержанием входных данных, а с другой – требованиями к технологическим процессам. Поскольку выходные данные отдельно взятого процесса служат исходными данными для других процессов, требования по обеспечению качества данных должны определяться на уровне всей цепочки передачи данных и согласованным образом предъявляться ко всем ИТ-процессам, задействованным в их переработке.

Общий подход к повышению качества данных должен предусматривать реализацию классического цикла Шухарта – Деминга (рис. 15.9) в той или иной вариации. Будучи основанным на методологии точных наук, этот четырехфазный цикл задает модель решения задачи методом последовательных приближений: планирование – реализация – проверка – корректировка – планирование – ...

Усовершенствования внедряются через строго определенную последовательность шагов. Применительно к программе качества данных это подразумевает следующий алгоритм действий: состояние данных подлежит контролю на предмет соответствия стандартам; если стандарты не соблюдены, требуется доработка, которая начинается с поиска и выявления корневых причин несоответствия данных стандартам с переходом на фазы планирования и реализации мер по устранению первопричин несоответствий, которые могут быть обусловлены технологическими, методологическими, организационными и человеческим факторами. После внесения всех необходимых исправлений и работы над ошибками система управления качеством данных продолжает функционировать в режиме мониторинга систем и контроля текущих данных на предмет выявления возможных новых нарушений стандартов.



Рис. 15.9.

Цикл Шухарта — Деминга*

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

Внедрение цикла управления качеством данных для набора данных, который ранее не отслеживался в рамках вышеописанной модели непрерывного совершенствования, начинается с выявления данных, не соответствующих стандартам или нуждам потребителей, и проблемных данных или процессов, препятствующих успешному решению стоящих перед бизнесом задач. Таким образом, данные нужно проверять на соответствие не только стандартам качества по всем ключевым параметрам, но и всем

известным бизнес-требованиям. Далее нужно устанавливать корневые причины несоответствий, чтобы все заинтересованные стороны могли объективно и взвешенно оценить затратность устранения недоработок и уровень риска в случае их сохранения. Эта часть работы обычно осуществляется совместно с распорядителями данных и иными заинтересованными лицами.

На стадии планирования команда качества данных составляет список текущих задач и проблем, сортирует их по масштабности и приоритетности, оценивает и сравнивает различные варианты решений. План должен строиться на прочном фундаменте анализа корневых причин. Без знания первопричин и последствий имеющихся проблем невозможны ни анализ полезности или эффективности затрат, ни определение приоритетов, а без этого ни о каком планировании говорить не приходится.

На стадии реализации команда качества данных руководит работами по устранению корневых причин имеющихся проблем, параллельно планируя показатели и средства последующего контрольного мониторинга данных. Если корневые причины носят нетехнический характер, команда качества данных совместно с владельцами процессов прорабатывают возможные процедурные изменения и порядок их осуществления. В случае проблем технического характера команда качества данных совместно с соответствующими инженерно-техническими службами обеспечивает надлежащую реализацию требуемых технических изменений и проверяет полученные результаты на предмет возможных ошибок.

На стадии проверки осуществляется активный мониторинг качества данных по заданным параметрам соответствия требованиям. Пока данные стабильно укладываются в контрольные допуски, дополнительных действий не требуется, а процессы считаются контролируемыми и соответствующими бизнес-требованиям. Но как только качество данных опускается ниже допустимого порогового уровня, необходимо принимать дополнительные меры по возвращению ситуации к норме.

Стадия корректировки включает работы по оперативному устранению текущих проблем с данными по мере их выявления системами контроля качества. Как только объем или характер текущих проблем выходят за рамки таких возможностей, цикл возобновляется и начинается поиск первопричин, а затем – проработка возможных решений.

Непрерывность обеспечения качества данных достигается за счет перезапуска цикла управления качеством данных в случае возникновения любой из перечисленных ниже ситуаций:

- выход текущих результатов измерений контрольных показателей за пределы допусков;
- появление новых наборов данных;
- изменение действующих или появление дополнительных требований к имеющимся наборам данных;
- изменение бизнес-правил, стандартов или ожиданий.

Сделать наборы данных правильными изначально – дешевле, чем исправлять неправильные наборы данных. Встроить процессы управления качеством данных в процессы оперативного управления данными с самого начала – на порядок дешевле, чем последующее исправление. Обеспечивать стабильно высокое качество данных на протяжении всего их жизненного цикла – менее рискованно, чем пытаться повышать качество данных в рамках существующих процессов. Перестройки на ходу достаточно тяжело бьют по организации. Определение критериев качества данных до начала планирования нового процесса или системы – признак зрелости организации в области управления данными и отличное средство укрепления административной дисциплины и налаживания плодотворного сотрудничества между функциональными подразделениями^{518, 519}.

⁵¹⁸ DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵¹⁹ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

УПРАВЛЕНИЕ КАЧЕСТВОМ ДАННЫХ

Определение: планирование, организация и контроль выполнения работ по применению стандартных методов управления качеством к данным с целью обеспечения их пригодности к использованию

Цели:

- 1. Разработка согласованного подхода к обеспечению соответствия данных потребностям потребителей
- 2. Определение стандартов и спецификаций контроля качества данных на протяжении их жизненного цикла
- 3. Разработка и внедрение процессов измерения, мониторинга и учета показателей качества данных
- 4. Выявление, изыскание и реализация возможностей для повышения качества данных посредством совершенствования систем и процессов

БИЗНЕС-ДРАЙВЕРЫ

Входные материалы • Политика и стандарты управления данными • Ожидаемые уровни качества данных • Бизнес-требования • Бизнес-правила • Требования к данным • Бизнес-метаданные • Технические метаданные • Источники и хранилища данных • Происхождение данных	Проводимые работы 1. Определение данных высокого качества (П) 2. Определение стратегии качества данных (П) 3. Определение объема работ по первоначальной оценке (П) • Выявление критически важных данных • Выявление действующих правил 4. Первичная оценка качества данных (П) • Выявление и приоритизация проблем • Анализ корневых причин проблем 5. Определение приоритетных мер (П) • Приоритетные меры с точки зрения бизнеса • Планы предупреждения и устранения ошибок • Утверждение планов работ 6. Разработка и внедрение операционных процедур обеспечения качества данных (Р, О) • Разработка операционных процедур DQ • Исправление данных • Мониторинг качества данных • Формирование отчетности в области DQ	Результаты • Стратегия и рамочная модель DQ • Организационная структура программы DQ • Результаты профилирования данных • Рекомендации по устранению корневых причин найденных проблем • Процедуры DQ • Текущие отчеты по DQ • Сводные отчеты по DQ для представления руководству • Соглашения об уровне обслуживания в области DQ • Политики, правила и методики DQ
Поставщики • Менеджмент бизнес-подразделений • Эксперты в предметных областях • Проектировщики архитектуры данных • Разработчики моделей данных • ИТ-специалисты • Распорядители данных • Бизнес-аналитики	Участники • CDO • Аналитики качества данных • Распорядители данных • Ответственные за управление данными • Аналитики данных • Администраторы баз данных • Профессионалы в области управления данными • Менеджеры DQ • ИТ-специалисты • Специалисты по интеграции данных • Команда по обеспечению нормативно-правового соответствия	Потребители • Бизнес-подразделения как потребители данных • Распорядители данных • Профессионалы в области управления данными • ИТ-специалисты • Работники интеллектуального труда • Органы руководства данными • Организации-партнеры • Центры компетенций

ТЕХНИЧЕСКИЕ ДРАЙВЕРЫ

Методы • Выборочная проверка подмножеств различных данных	Инструменты • Инструменты профилирования • Инструменты для создания запросов	Метрики • Показатели соблюдения правил и процедур • Результаты измерения показателей

* DAMA. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

15.3.5. Контекстная диаграмма области знаний и уровни зрелости функции «Качество данных»

Контекстная диаграмма области знаний «Качество данных» представлена на рисунке 15.10.

На рисунке 15.11 отражено распределение деятельности в области управления качеством данных по этапам их жизненного цикла.

На рисунке 15.12 представлены обобщенные характеристики уровней зрелости функции «Качество данных». Для каждого уровня, помимо характеристик зрелости практик управления качеством данных в целом, отдельно приведены характеристики деятельности по измерению качества.



Рис. 15.11.

Распределение деятельности в области управления качеством данных по этапам их жизненного цикла*

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

Практика обеспечения качества данных выступает в роли ключевого звена организационной стратегии и рассматривается как существенный фактор успешной реализации каждого проекта. Предпринимаются активные шаги по улучшению качества данных с целью повышения их ценности для всех заинтересованных сторон.

Измерение качества: отчеты о качестве данных широко распространяются среди заинтересованных сторон. Результаты измерения качества используются для определения возможностей по совершенствованию систем и бизнес-процессов. Об итогах проводимых улучшений сообщается. Деятельность по подготовке отчетов и оповещению полностью автоматизирована.

УРОВЕНЬ 5 — ОПТИМИЗИРОВАННЫЙ

Практики в области обеспечения качества данных измеряются и контролируются согласно общеорганизационным стандартам, при этом все бизнес-функции отчитываются о соответствии (или несоответствии) управляющему комитету (группе) по руководству данными.

Измерение качества: владельцы систем и бизнес-процессов обязаны измерять качество своих данных и сообщать о результатах, чтобы потребители данных имели ясное представление об их качестве. Процесс измерения полностью автоматизирован.

УРОВЕНЬ 4 — УПРАВЛЯЕМЫЙ

Определены общеорганизационные политики и стандарты, устанавливающие стратегическую линию в отношении применения практики обеспечения качества данных по всем бизнес-функциям. Бизнес-функции устанавливают свои локальные практики на основе этих стандартов.

Измерение качества: определены стандарты для измерения качества. Приняты стандартные подходы к автоматизации.

УРОВЕНЬ 3 — УСТАНОВЛЕННЫЙ

В рамках отдельных бизнес-функций или подразделений имеются свидетельства наличия документированных практик обеспечения качества данных, хотя нет никакой гарантии, что они являются достаточно всеобъемлющими или что существует какая-либо согласованность в действиях по их внедрению по подразделениям.

Измерение качества: сотрудники изучили подходы к оценке качества данных и разрабатывают способы измерения. Операции по-прежнему в основном выполняются вручную, но некоторые команды протестировали средства автоматизации.

УРОВЕНЬ 2 — ПОВТОРЯЕМЫЙ

В организации нет никаких свидетельств применения практик обеспечения качества за рамками отдельных систем.

Измерение качества: отдельные сотрудники пытаются измерять качество данных, но это не входит непосредственно в их обязанности и у них нет определенной методики. Измерения производятся вручную.

УРОВЕНЬ 1 — НАЧАЛЬНЫЙ

* Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

* Smith P., Edge J., Parry S., Wilkinson, D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.



ПРАКТИЧЕСКИЙ ПРИМЕР

Вернемся к герою рубрики «Практический пример» – компании «Телеком Дубль». Говоря о базовой поддержке жизненного цикла данных, следует заметить, что особую роль здесь играет безопасность. Хороший пример – обеспечение требований Федерального закона № 152 «О персональных данных». Имея в арсенале централизованную модель данных с описанием клиента компании, «Телеком Дубль» может определять системы, которые используют сведения о клиентах, маркировать «чувствительные» данные и, исходя из этого, ограничить доступ к ним со стороны тех систем, для которых не предусмотрена работа с защищенной информацией.

В части управления метаданными компания инициировала важные работы по созданию бизнес-гlossария и внедрению каталога технических метаданных.

Наконец, был сделан важнейший шаг по обеспечению базовой поддержки жизненного цикла информационных ресурсов «Телеком Дубль» – проведение мероприятий по разработке стратегии и рамочной модели системы обеспечения качества данных.

Литература к главе 15

- Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».
- Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных».
- ГОСТ Р ИСО/МЭК 11179-1-2010. Информационная технология. Регистры метаданных (РМД). Часть 1. Основные положения.
- ГОСТ Р 56214-2014/ISO/TS8000-1:2011. Качество данных. Часть 1. Обзор.
- ГОСТ Р ИСО 8000-2-2019. Качество данных. Часть 2. Словарь.
- ГОСТ Р ИСО/МЭК 27001-2021. Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования.
- ГОСТ Р ИСО/МЭК 27002-2021. Информационные технологии. Методы и средства обеспечения безопасности. Свод норм и правил применения мер обеспечения информационной безопасности.
- ГОСТ Р ИСО/МЭК 27005-2010. Информационная технология. Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности.
- ГОСТ ISO/IEC 29100-2021. Информационные технологии. Методы и средства обеспечения безопасности. Основы защиты персональных данных.
- Smith P., Edge J., Parry S., Wilkinson, D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.
- Van Gils B. Data Management: a Gentle Introduction: Balancing Theory and Practice. Van Haren Publishing, 2020.

Глава 16. Организационные аспекты управления данными

Описывая в главе 9 рамочную структуру функций управления данными в привязке к их жизненному циклу (рис. 9.4), мы заметили, что она, по сути, отражает модель управления цепочками поставок данных. В главах 10–15 были рассмотрены все функциональные блоки модели за исключением финального – «Использование данных и расширение возможностей применения». Прежде чем перейти к обсуждению этого ключевого с точки зрения извлечения ценности блока, в интересах которого и формируется вся цепочка поставок данных, стоит еще раз более подробно поговорить об организационных аспектах ее

формирования. На важность этой темы мы уже обращали внимание, когда рассматривали «Сегодняшнюю повестку дня для совета директоров» (см. главу 7) и вопросы руководства данными (см. главу 10).

В главе 10 уже обсуждалась обобщенная организационная система руководства и управления данными (рис. 10.5). Ее можно рассматривать как совокупность двух взаимодействующих подсистем: организационной системы руководства данными (Data Governance Organization, DGO) и организационной системы управления данными (Data Management Organization, DMO). В данной главе описаны общие принципы, которые следует учитывать при их построении. Принципы относятся как к DGO, так и к DMO, поскольку руководство данными задает направления и предоставляет бизнес-контекст для работ, выполняемых системой управления данными. Для этих систем невозможно предложить идеальную структуру. В то время как общие принципы распространяются на каждую из них, детали будут сильно зависеть от отраслевых факторов и корпоративной культуры самой организации⁵²⁰.

В заключение главы рассмотрены основные шаги, которые являются важнейшими при инициировании в организации деятельности по совершенствованию управления данными.

16.1. Операционные модели руководства данными и управления данными

Критически важный шаг на пути создания организационной системы (как руководства, так и управления данными) – определение наиболее подходящей операционной модели. Операционная модель служит рамочной структурой для определения ролей, обязанностей и процессов принятия решений. Она описывает порядок взаимодействия людей и функций.

Надежная операционная модель помогает наладить механизмы подотчетности, поскольку в ней представлены все необходимые функции. Она способствует развитию коммуникаций и обеспечивает поддержку процесса разрешения проблемных ситуаций. Формируя базис для организационной структуры, операционная модель не задает ее фиксированную схему. Речь идет не о составлении штатного расписания, а об описании взаимосвязей между составными частями организационной системы.

16.1.1. Операционные модели руководства данными

При построении операционной модели DG рекомендуется учитывать нижеследующие аспекты⁵²¹.

- **Ценность данных для организации.** Если организация занимается продажей данных, значительное влияние DG на бизнес очевидно. Организации, использующие данные в качестве ключевого товара (например, агрегаторы), нуждаются в операционной модели, отражающей роль данных в достижении успеха. В то же время в организациях, где данные используются только в качестве «смазочного материала» для осуществления операционной деятельности, программа DG будет реализована менее строго.

- **Бизнес-модель.** Бизнес-модель может быть централизованной или децентрализованной, локальной или международной и т. п. – все факторы подобного рода оказывают влияние на информационные потребности бизнеса и, как следствие, на определение операционной модели DG. В проекте целевой операционной рамочной структуры должны быть отражены специфические для данной организации связи с ИТ-стратегией, архитектурой данных и решениями по интеграции приложений.

- **Культурные факторы.** Такие как дисциплинированность и адаптируемость к изменениям. Некоторые организации будут сопротивляться навязыванию руководства посредством внедрения политик и принципов. В таком случае стратегия руководства данными потребует для разъяснения преимуществ операционной модели DG, которая при этом должна вписаться в сложившуюся организационную культуру (с учетом последующего проведения поэтапных изменений).

- **Влияние регламентации.** В организациях, деятельность которых сильно зарегламентирована, менталитет сотрудников существенно отличается от менталитета сотрудников в относительно свободных

⁵²⁰ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵²¹ Там же.

организациях. Соответственно, для организаций первого типа должна отличаться и операционная модель DG. В частности, в ней могут быть предусмотрены связи с группой управления рисками, а также с юридической службой.

Часто в операционной модели выделяются несколько ступеней, чтобы обеспечивать решение вопросов и проблем на разных уровнях управления – локальном, дивизиональном и корпоративном (рис. 10.5). Такой подход подразумевает четкое определение сфер ответственности: например, кто отвечает за проведение работ по распоряжению данными или кто является владельцами данных. Операционная модель также определяет порядок взаимодействия между организационной системой руководства данными и ответственными за различные проекты или инициативы по управлению данными, порядок согласования и проведения мероприятий по управлению изменениями, необходимыми для реализации этой новой программы, а также модель разрешения проблемных вопросов в рамках руководства данными. Рисунок 16.1 описывает пример операционной модели DG.



COO (Chief Operations Officer) — главный операционный директор, CFO (Chief Financial Officer) — финансовый директор; DGC (Data Governance Council) — совет по руководству данными, ERP (Enterprise Resource Planning) — планирование ресурсов предприятия

Рис. 16.1.

Пример операционной модели DG *

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Приведенный пример иллюстративен.Arteфакты подобного рода должны быть адаптированы к потребностям конкретной организации.

Одни операционные модели предусматривают централизованное управление, другие –

распределенное. Основные типы моделей в зависимости от степени централизации представлены на рисунке 16.2.

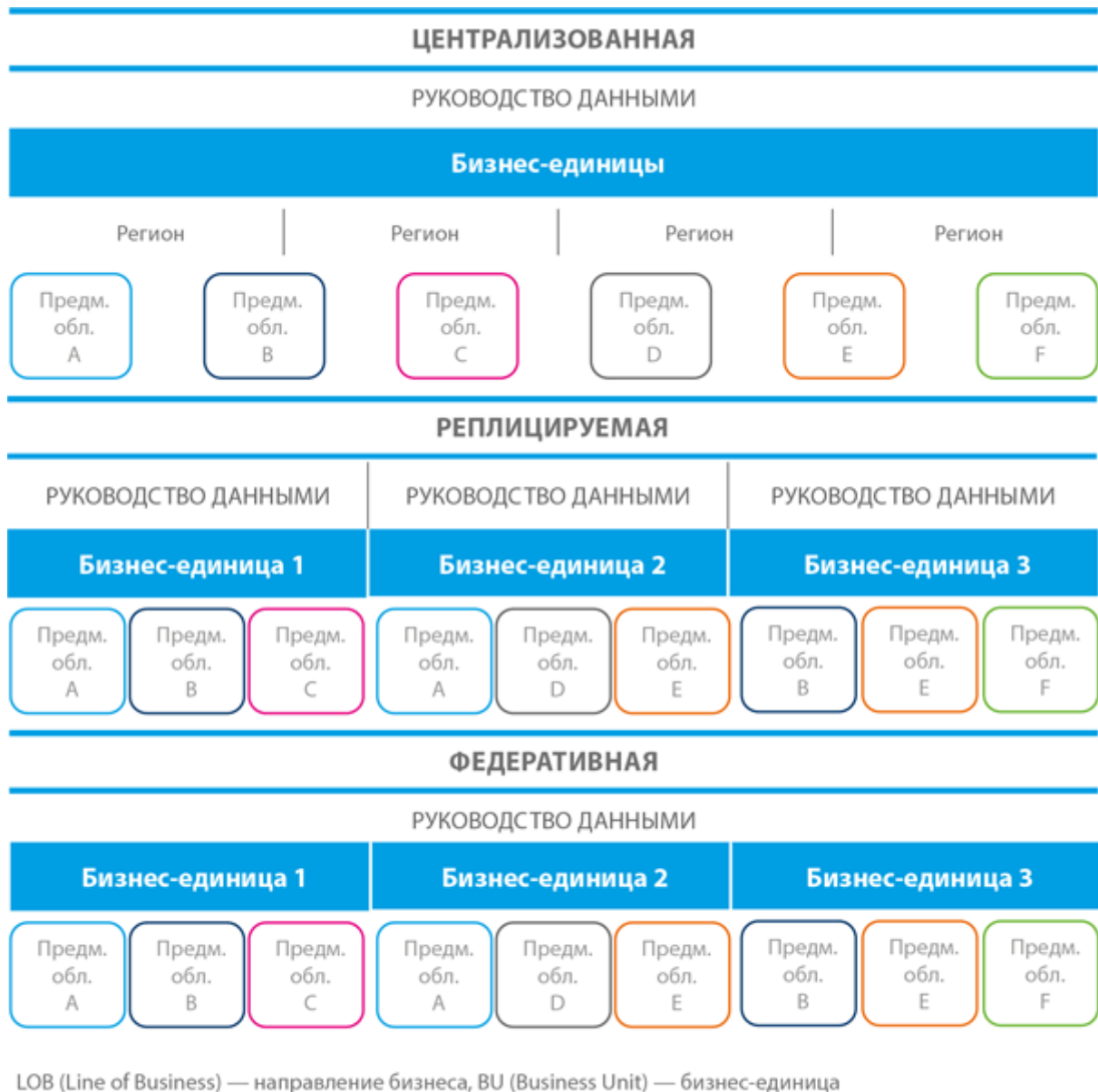


Рис. 16.2.

Основные типы операционных моделей DG в зависимости от степени централизации

При централизованной модели одна организационная система руководства данными контролирует работы по всем предметным областям. В реплицируемой модели одни и те же операционная модель и стандарты DG воспроизводятся в каждой бизнес-единице. Наконец, при федеративной модели одна организационная система руководства данными координирует деятельность нескольких бизнес-единиц с целью обеспечения согласованности определений и стандартов.

16.1.2. Операционные модели управления данными

Операционные модели управления данными можно разделить на следующие типы:

- децентрализованная;
- сетевая;
- централизованная;
- гибридная;
- федеративная.

Большинство организаций начинают с децентрализованной модели и лишь через какое-то время осознают необходимость перехода к более формальной организационной системе управления данными. Обратив внимание на тот факт, что упорядочение управления данными позитивно влияет на их качество, организация может приступить к формализации ответственности в сфере управления данными, начав движение к сетевой, а затем и централизованной модели⁵²².

Децентрализованная операционная модель

В рамках децентрализованной модели ответственность за управление данными распределяется по различным направлениям бизнеса и деятельности в области ИТ (рис. 16.3).

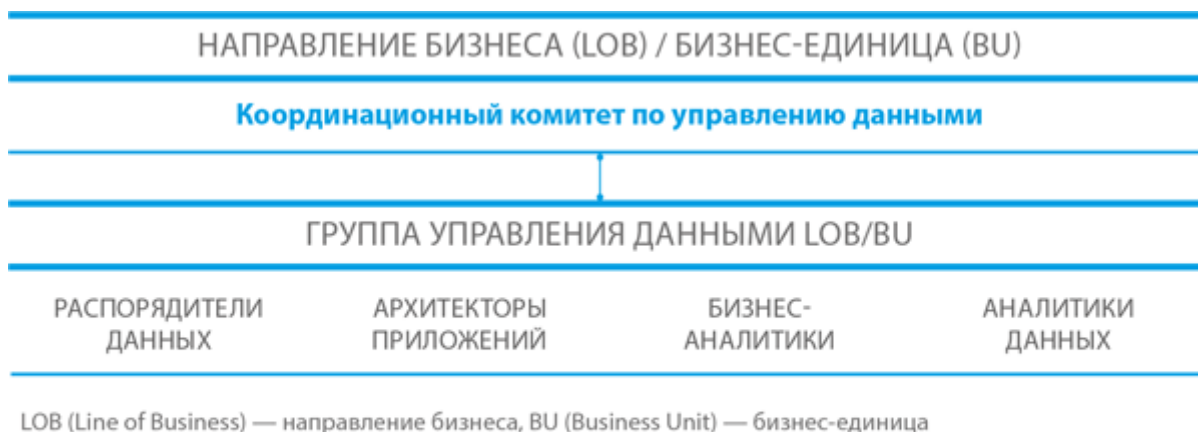


Рис. 16.3.

Децентрализованная операционная модель ДМ*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Любое сотрудничество возможно только через комитеты; единый ответственный отсутствует. Многие программы управления данными начинаются по инициативе снизу, направленной на то, чтобы хоть как-то упорядочить практику управления данными в масштабах организации, а потому носят по определению децентрализованный характер.

Сетевая операционная модель

Децентрализованную модель можно сделать более строгой и формализованной посредством ее дополнения задокументированным распределением ролей и обязанностей, как правило описываемым через матрицу RACI⁵²³. Такую модель принято называть сетевой, поскольку она работает по принципу использования горизонтальных связей между людьми, участвующими в процессе в различных ролях, и может быть отражена в виде сети (рис. 16.4).

⁵²² Там же.

⁵²³ RACI (сокр. от англ. Responsible / Accountable / Consulted / Informed) используется для назначения ответственного исполнителя (R), куратора (A), консультантов (C) и информируемых (I) в рамках проекта или направления работы.

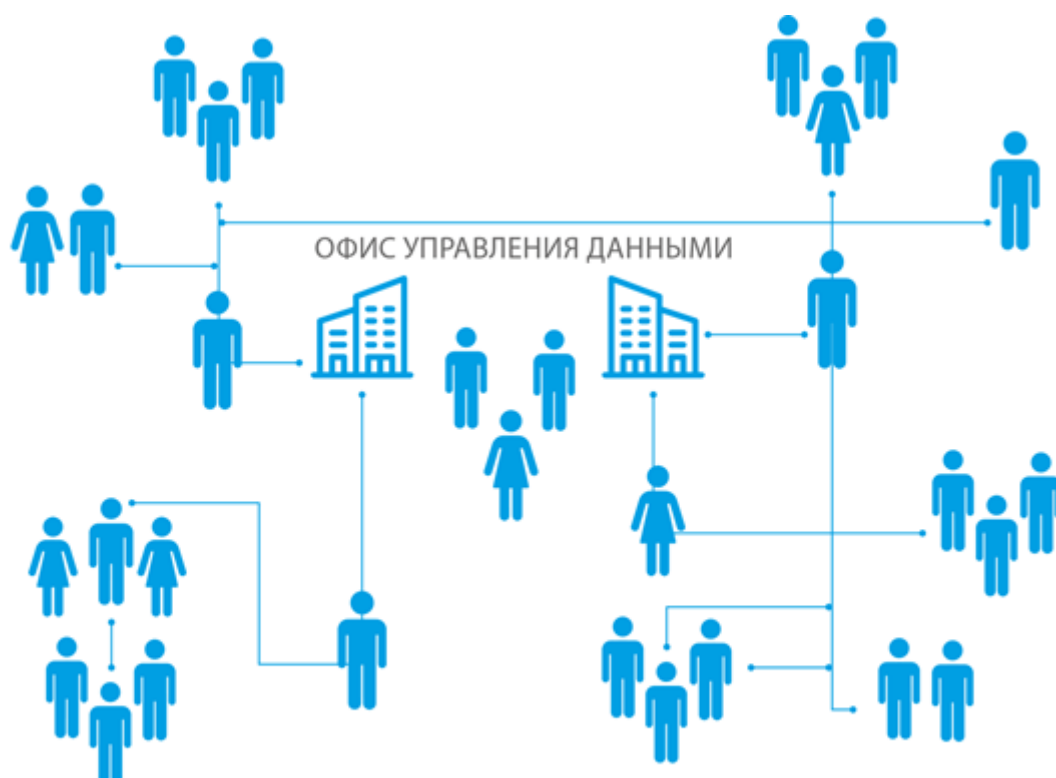


Рис. 16.4.

Сетевая операционная модель ДМ*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

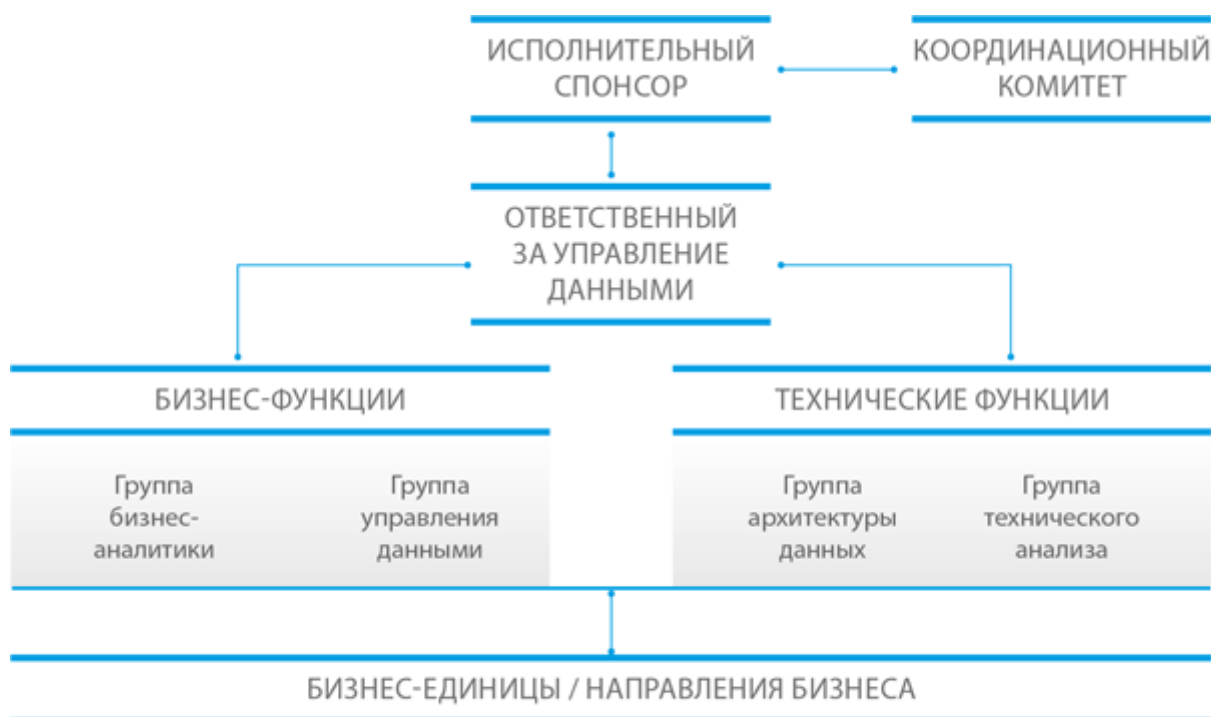


Рис. 16.5.

Централизованная операционная модель ДМ*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Централизованная операционная модель

Самая формализованная и зрелая централизованная операционная модель управления данными (рис. 16.5). В ней все упорядочено и входит в сферу влияния организационной системы управления данными. Все участники процессов руководства и управления данными отчитываются непосредственно перед главным руководителем по вопросам управления данными, отвечающим за руководство и распоряжение данными, управление метаданными, управление качеством данных, управление справочными и основными данными, архитектуру данных, а также бизнес-аналитику.

Гибридная операционная модель

Гибридная операционная модель совмещает преимущества централизованной и децентрализованной моделей (рис. 16.6). В рамках гибридной модели головной центр компетенций в области управления данными координирует работу децентрализованных групп управления данными бизнес-единиц, стратегические направления совершенствования обычно определяются координационным комитетом, в котором представлены все ключевые направления бизнеса, а тактические вопросы решаются на уровне рабочих групп, создаваемых внутри бизнес-единиц и руководствующихся методическими рекомендациями центра компетенций.

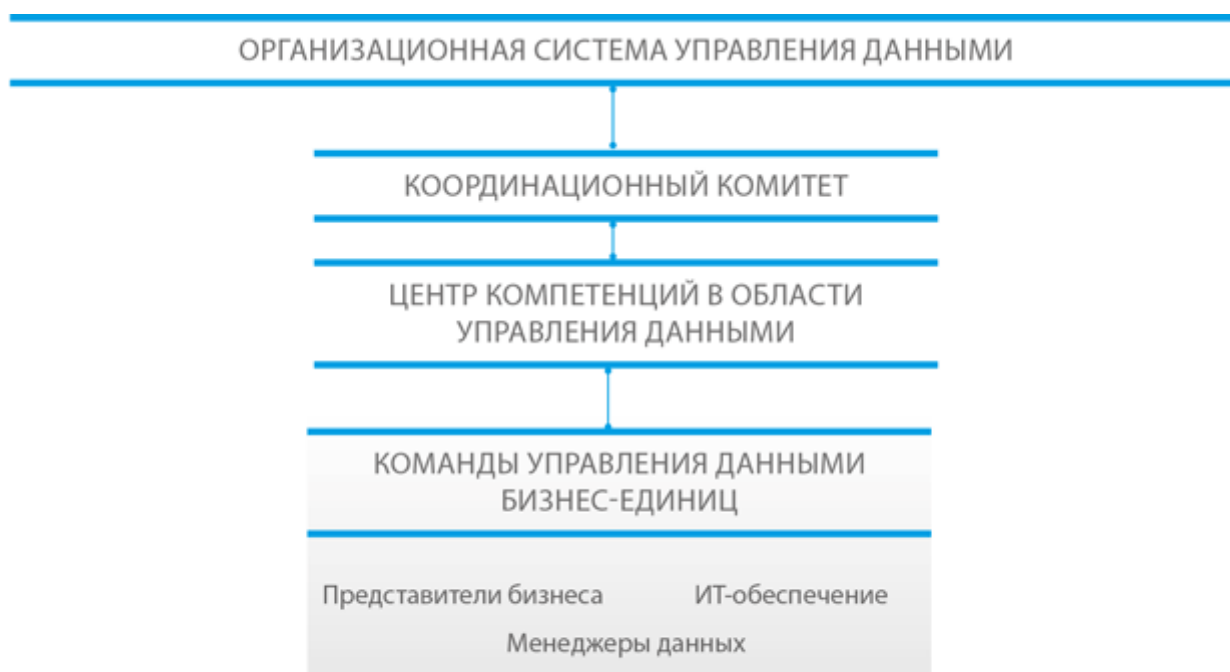


Рис. 16.6.

Гибридная операционная модель DM*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

Гибридная модель оставляет часть ролевых функций децентрализованными. Например, проектировщики архитектуры данных организационно могут оставаться в составе группы управления корпоративной архитектурой, а функции обеспечения качества данных – в ведении руководства направлениями бизнеса, у каждого из которых будет собственная команда качества данных. Какие роли централизовать, а какие оставить децентрализованными, зависит от множества факторов, обуславливаемых преимущественно организационной культурой.

Федеративная операционная модель

Разновидностью гибридной операционной модели является федеративная модель,

предусматривающая многоуровневую централизацию. Обычно такой подход используется лишь в крупнейших транснациональных корпорациях и глобальных организациях. Федеративная модель представляет собой множество различных параллельно реализованных в разных регионах или направлениях бизнеса гибридных операционных моделей (рис. 16.7).

Федеративная модель обеспечивает децентрализованное исполнение централизованно вырабатываемых стратегических планов. Для крупных многопрофильных организаций она, возможно, единственная работоспособная модель организационной системы управления данными.



Рис. 16.7.

Федеративная операционная модель DM*

* DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

16.1.3. Взаимодействие операционных моделей

Руководство данными – это организационная система для выработки стратегии, целей и задач, политики и правил эффективного управления данными. Сюда также относятся процессы, регламенты, организационные подсистемы и технологии, необходимые для обеспечения доступности, годности, целостности, согласованности, достоверности и защищенности данных. Поскольку программа руководства данными состоит из хитросплетения рабочих стратегий, стандартов, политик и коммуникаций, так или иначе относящихся к данным, она работает в тесной синергетической связи с управлением данными, определяя рамки выстраивания процессов управления в соответствии с приоритетами бизнеса и заинтересованных сторон.

При централизованной модели управления данными организационная система управления данными (в лице офиса руководства данными) либо подчинена организационной системе управления данными, либо наоборот. Когда программа управления данными ориентирована прежде всего на выработку политики и правил управления данными как ценным активом, логично осуществлять ее под общим руководством

офиса руководства данными, сделав оргсистему управления данными подчиненной или подотчетной (или выстраивать DMO в формате матричного подчинения офису руководства данными по направлениям работы). Такой подход часто наблюдается в сильно зарегулированных организациях, где главное – соблюдение правил и строгая отчетность.

Но и в самых децентрализованных вариантах моделей управления данными необходимо тесное партнерство между DGO и DMO с целью четкого разграничения и согласования функций разработки политик, правил и инструкций в сфере управления данными (DGO) и их реализации (DMO). Это две стороны одной медали и два основных компонента, необходимые для получения ценных данных. Таким образом, можно считать руководство данными методом упорядочения управления данными или, образно выражаясь, функцией штабного планирования боевых операций.

Необходимо четко понимать синергию взаимодействия двух операционных моделей и согласовывать роли, обязанности и сферы ответственности с целью одновременного выполнения указаний и инструкций по руководству данными и обеспечения эффективности оперативного управления данными. К участию в рабочей группе по руководству данными целесообразно привлекать представителей организационной системы управления данными, а DMO может исполнять роль уполномоченного исполнительного звена и осуществлять надзорные функции от имени и по поручению представителей руководства данными при поддержке его административного ресурса⁵²⁴.

16.2. Роли, относящиеся к руководству и управлению данными

Роли, связанные с руководством и управлением данными, могут определяться как на функциональном (организационные роли), так и на индивидуальном (индивидуальные роли) уровне. При этом и названия ролей, и их относительная значимость и нужность могут варьироваться от организации к организации⁵²⁵.

16.2.1. Организационные роли

Организационные системы управления данными, относящиеся к блоку ИТ, предлагают широкий спектр сервисов, начиная с проектирования ИТ-инфраструктуры и архитектуры данных/приложений и заканчивая поставкой под ключ и администрированием СУБД.

При централизованной сервисной организационной системе управления данными все внимание уделяется управлению данными как таковому. В такую команду могут входить исполнительный директор по управлению данными, подотчетные ему менеджеры, архитекторы и аналитики данных, эксперты по качеству данных, администраторы баз данных, администраторы информационной безопасности, специалисты по метаданным, моделированию и администрированию данных, проектировщики архитектуры хранилищ и интеграции данных, аналитики BI – иными словами, представители всех групп, имеющих прямое или косвенное отношение к сбору, обработке и хранению данных.

При частично распределенной сервисной организационной системе управления данными (федеративной модели) имеется ряд параллельно функционирующих ИТ-подразделений, каждое из которых отвечает за обеспечение своего среза работ по управлению данными. Как правило, чем крупнее организация, тем больше децентрализуются ИТ-функции. Например, каждой бизнес-функции может быть придана собственная группа разработчиков ПО. Возможен и гибридный подход, например: разработчики приложений у каждой бизнес-функции имеют собственные, а функция администрирования БД остается централизованной.

Бизнес-функции, обеспечивающие управление данными, чаще всего ассоциируются с командами по руководству данными или корпоративному управлению информацией. Например, распорядители данных часто входят в состав организационной системы по руководству данными, что облегчает работу органов руководства данными, в частности таких, как совет по руководству данными.

⁵²⁴ Там же.

⁵²⁵ Там же.

16.2.2. Индивидуальные роли

Индивидуальные роли могут определяться в рамках бизнес– или ИТ-функций. Часть ролей так или иначе будет иметь смешанный (гибридный) характер, поскольку требует от исполнителей знания как информационных систем, так и бизнес-процессов.

Руководящие роли

Исполнительное руководство управления данными в зависимости от ситуации может представлять как сферу ИТ, так и сферу бизнеса. Должности директора по ИТ (CIO) и технического директора (CTO) уже получили повсеместное распространение, а в последние годы и позиция директора по данным (CDO) начала активно приживаться в практике корпоративного управления.

Бизнес-роли

Бизнес-роли преимущественно относятся к функциям руководства данными. Прежде всего, речь идет о распорядителях данных по направлениям деятельности. Как правило, это признанные эксперты в предметных областях, которые отвечают за метаданные и качество данных как по отдельным бизнес-сущностям, так и по предметным областям, и в целом по базам данных. Роли распорядителей данных могут серьезно различаться в зависимости от специфики и приоритетов организации. Часто они отвечают за обеспечение точности определений бизнес-терминов и областей значений различных данных в соответствующих предметных областях. Но со временем во многих организациях задействуют этих же распорядителей для определения критериев качества данных, бизнес-правил и атрибутов данных в профильной области, выявления и содействия разрешению проблемных вопросов. Они же незаменимы при определении стандартов, правил и процедур. Распорядители данных могут функционировать на разных уровнях – предприятия, бизнес-подразделения или функции управления. Роли этих людей могут быть как формализованными на уровне должностей (например, ответственный за конкретные данные), так и совершенно неформальными (люди просто заботятся о том, чтобы с данными по их линии работы все было в порядке, как бы ни называлась их должность).

В зависимости от сложности организации и целей ее программы DG официально назначенные распорядители данных могут различаться по своей позиции в организации, направлению работы или по обоим указанным признакам.

- **Главные распорядители данных** (Chief Data Stewards) могут возглавлять органы руководства данными вместо CDO или выступать в качестве CDO в виртуальной (основанной на комитетах) или распределенной организационной системе руководства данными. Они также могут быть исполнительными спонсорами (Executive Sponsors).

- **Исполнительные распорядители данных** (Executive Data Stewards) – это старшие руководители, входящие в состав совета по руководству данными (DGC).

- **Распорядители корпоративных данных** (Enterprise Data Stewards) осуществляют надзор (oversight) за отдельными областями (domain) данных предприятия в процессе выполнения всех связанных с этими областями бизнес-функций.

- **Распорядители бизнес-данных** (Business Data Stewards) – это бизнес-специалисты, чаще всего признанные эксперты в той или иной предметной области, ответственные за соответствующее подмножество данных. Они работают с заинтересованными лицами (stakeholders) в части определения и контроля данных.

- **Владелец данных** (Data Owner) – это распорядитель бизнес-данных, который обладает подтвержденными полномочиями на утверждение решений, касающихся его области данных.

- **Технические распорядители данных** (Technical Data Stewards) – это ИТ-специалисты, работающие в одной из областей знаний управления данными. Среди них – специалисты по интеграции данных, администраторы баз данных, специалисты по бизнес-аналитике, аналитики качества данных или администраторы метаданных.

- **Координирующие распорядители данных** (Coordinating Data Stewards) возглавляют и представляют команды распорядителей бизнес-данных и технических распорядителей данных в обсуждениях как на уровне команд, так и с участием исполнительных распорядителей данных. Координирующие распорядители данных особенно важны в крупных организациях.

Помимо формальных распорядителей данных, свой важный вклад в организацию управления данными вносят бизнес-аналитики и разработчики моделей бизнес-процессов, поскольку без их участия попросту невозможно обеспечить соответствие моделируемых процессов реальным и, как следствие, пригодность накапливаемых данных для использования их ниже по информационно-технологическому потоку.

Вносят свой посильный вклад в общеорганизационное управление данными и другие категории компетентных сотрудников со стороны бизнеса: например, потребители аналитики, оставляющие замечания по поводу публикуемых организацией данных, которые способствуют всестороннему совершенствованию управления данными.

ИТ-роли

В управлении данными задействуются специалисты по ИТ всевозможных профилей и уровней, включая архитекторов, разработчиков и системных администраторов баз данных и приложений, многочисленный технический персонал. Перечислим самые распространенные роли.

- **Архитектор данных:** старший аналитик, отвечающий за архитектуру и интеграцию данных на уровне предприятия или функционального подразделения. В зависимости от профиля архитекторы данных могут специализироваться на построении хранилищ данных, витрин данных и процессов интеграции.

- **Разработчик модели данных** отвечает за выявление и структурное описание (моделирование) требований к данным, определение объектов и элементов данных, а также связей между ними, бизнес-правил, требований к качеству данных и в целом логических и физических моделей данных.

- **Администратор модели данных** отвечает за управление версиями модели данных и их своевременное и согласованное обновление.

- **Администратор базы данных** отвечает за своевременное получение и обработку массивов входящих данных, а также технологическое обеспечение их доступности.

- **Администратор информационной безопасности** отвечает за контроль доступа к данным в зависимости от уровня защиты данных и прав доступа, имеющихся у запрашивающих доступ сторон.

- **Архитектор интеграции данных** отвечает за принципиальное обеспечение совместимости и качества данных на уровне предприятия.

- **Специалист по интеграции данных** проектирует, разрабатывает и внедряет системы интеграции (копирования, извлечения, преобразования, загрузки) массивов необходимых данных в требуемых режимах (пакетной или потоковой обработки).

- **Разработчик аналитических или статистических отчетов** занимается созданием программных средств генерирования отчетов в согласованных форматах.

- **Архитектор приложений** отвечает за интеграцию прикладного ПО с системами управления данными.

- **Технический архитектор** координирует работы по интеграции портфеля новых ИТ-проектов в существующую инфраструктуру.

- **Технический инженер** отвечает за изыскание и реализацию (в пределах своей компетенции) возможностей совершенствования ИТ-инфраструктуры.

- **Администратор службы поддержки** отвечает за своевременный прием, обработку и отслеживание всех сигналов, а также контроль разрешения проблем, связанных с передачей или получением данных, работой информационных систем или ИТ-инфраструктуры.

- **ИТ-аудитор:** более чем желательно наличие внешнего или внутреннего независимого контроля соблюдения всех критериев приемлемости ИТ-обеспечения управления данными, включая контроль качества и обеспечение надлежащего уровня ИБ.

Гибридные роли

Гибридные роли требуют от исполнителей сплава технических навыков со знанием бизнеса. В зависимости от специфики организации такие специалисты могут формально относиться как к бизнес-подразделениям, так и к ИТ-службам.

- **Аналитик качества данных** отвечает за пригодность данных к использованию и текущий мониторинг состояния данных; участвует в анализе корневых причин выявленных проблем с данными и

вырабатывает рекомендации по реорганизации бизнес-процессов и совершенствованию ИТ-решений, направленные на устранение недостатков и повышение качества данных.

- **Специалист по метаданным** отвечает за их определение и интеграцию, управление метаданными и их своевременное обновление, исполняет функции администратора хранилищ метаданных.

- **Архитектор ВІ:** старший бизнес-аналитик, отвечающий за выбор и интеграцию приложений в пользовательской ВІ-среде.

- **Администратор ВІ** отвечает за эффективный доступ бизнес-пользователей к средствам и результатам бизнес-анализа.

- **Руководитель программы ВІ** координирует на корпоративном уровне ВІ-требования, инициативы и проекты, обеспечивает их интеграцию в комплексную программу приоритетных ВІ-исследований и осуществляет оперативное планирование ее реализации.

16.3. Ключевые первоначальные шаги

В заключение этой главы рассмотрим шаги, которые являются важнейшими при инициировании в организации деятельности по совершенствованию управления данными. К ним относятся:

- оценка текущего состояния;
- выработка понимания возможностей по улучшению с целью разработки дорожной карты реализации программы управления данными;
- инициирование программы управления организационными изменениями, поддерживающей выполнение дорожной карты⁵²⁶.

16.3.1. Оценка текущего состояния

Первый шаг к решению проблемы – формирование четкого представления о ее сути. Прежде чем определять какую-либо новую систему управления или попытаться улучшить существующую, важно понять текущее состояние составляющих ее элементов, особенно если они связаны с культурой, существующей операционной моделью и людьми. Хотя специфика требуемых изменений, направленных на улучшение управления данными, своя для различных организаций, оценка текущего состояния во всех случаях должна учитывать такие параметры:

1. **Роль данных в организации** : какие ключевые процессы управляются на основе данных? Как определяются и понимаются требования к данным? Насколько хорошо осознается роль, которую данные играют в стратегии организации? В какой степени организация осведомлена об издержках, связанных с некачественными данными?

2. **Культурные нормы в отношении данных:** существуют ли потенциальные культурные препятствия для внедрения или совершенствования структур управления данными и управления? Осведомлены ли владельцы бизнес-процессов о том, как их данные используются в связанных процессах?

3. **Практика руководства и управления данными:** как и кем выполняется работа, связанная с данными? Как и кем принимаются решения о данных?

4. **Как организована и выполняется работа:** какова связь между оперативной и проектной деятельностью? Какие существуют комитеты или элементы оргструктуры, которые могут поддерживать усилия по управлению данными? Какова операционная модель взаимодействия ИТ и бизнеса? Как финансируются проекты?

5. **Отношения подотчетности:** является ли организация централизованной или децентрализованной, иерархической или плоской? Насколько тесно взаимодействуют команды?

6. **Уровни квалификации:** каков уровень знаний в области данных и управления данными у экспертов в предметных областях и других заинтересованных сторон, от рядовых сотрудников до руководителей?

Оценка текущего состояния также должна включать оценку уровня удовлетворенности текущим

⁵²⁶ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

состоянием. Это позволит получить представление о потребностях и приоритетах организации в области управления данными. Например:

1. **Принятие решений:** располагает ли организация информацией, необходимой для принятия обоснованных и своевременных бизнес-решений?
2. **Отчетность:** уверена ли организация в своих отчетах о доходах и других важных данных?
3. **Ключевые показатели эффективности:** насколько хорошо организация отслеживает ключевые показатели эффективности?
4. **Нормативно-правовое соответствие:** соблюдает ли организация требования нормативно-правовых актов, касающихся управления данными?

Наиболее эффективный способ проведения такой оценки – использование проверенной модели зрелости управления данными, которая позволит получить представление как об уровне зрелости организации по сравнению с другими организациями, так и о дальнейших шагах по его повышению.

Основная цель оценки текущего состояния – понять базовый уровень организации, являющийся отправной точкой для планирования улучшений. Точная оценка важнее, чем высокий оценочный балл. Формальная оценка зрелости управления данными позволяет соотнести уровень организации со шкалой зрелости, уточняя конкретные сильные и слабые стороны важнейших видов деятельности по управлению данными. Это помогает организации определять, расставлять приоритеты и реализовывать возможности по улучшению.

Выполняя основную задачу, проведение оценки зрелости может оказать положительное влияние на организационную культуру. Она способствует:

- информированию заинтересованных сторон о концепциях, принципах и методах управления данными;
- разъяснению ролей и обязанностей заинтересованных сторон в отношении данных;
- осознанию необходимости управления данными как важнейшим активом;
- расширению признания деятельности по управлению данными в масштабах всей организации;
- улучшению сотрудничества, необходимого для эффективного управления данными.

Основываясь на результатах оценки, организация может усовершенствовать свою программу управления данными с целью лучшей поддержки выполнения операционных и стратегических задач. Как правило, разработка программ управления данными редко начинается с выработки корпоративного представления о данных, чаще такие программы фокусируются на мероприятиях в отношении отдельных подразделений. Оценка текущего состояния может помочь организации разработать целостное видение, поддерживающее общеорганизационную стратегию. Она позволяет организации уточнить приоритеты, четко сформулировать цели и разработать комплексный план улучшения.

16.3.2. Использование результатов для планирования улучшений

Оценка текущего состояния позволяет определить, что работает хорошо, что недостаточно хорошо и где организация имеет пробелы. Основываясь на полученных данных, организация может разработать дорожную карту, нацеленную на:

- совершенствование по наиболее важным (обеспечивающим наибольшую выгоду) направлениям, относящимся к процессам, методам, ресурсам и средствам автоматизации;
- обеспечение возможностей, которые соответствуют бизнес-стратегии;
- поддержку процессов руководства, которые необходимы для периодической оценки прогресса организации, основанной на характеристиках, заложенных в модель оценки зрелости.

16.3.3. Инициирование программы управления организационными изменениями, поддерживающей выполнение дорожной карты

Большинство организаций, которые стремятся улучшить свои методы управления данными, находятся в середине шкалы зрелости способностей. Они не имеют ни 0, ни 5 баллов по шкале зрелости. Это означает, что почти всем из них необходимо совершенствовать свою практику.

Для большинства организаций совершенствование методов управления данными требует изменения подходов к обеспечению взаимодействия сотрудников и их взглядов на роль данных в организации. Кроме

того, должны быть изменены представления об использовании данных и внедрении технологий для поддержки организационных процессов. Успешная практика управления данными требует среди прочего:

- обучения управлению по горизонтали путем выравнивания ответственности по цепочке создания ценности информации;
- смещения акцента с вертикальной (изолированной) подотчетности на совместное распоряжение информацией;
- превращения мероприятий по обеспечению качества информации из частных бизнес-задач или задач блока ИТ в ключевую деятельность по созданию ценности для организации;
- перехода от представлений о качестве информации, связанных в первую очередь с очисткой данных и разработкой соответствующей системы показателей качества, к представлениям о более фундаментальной организационной способности, связанной со встраиванием качества в процессы;
- внедрения процессов по измерению величины затрат, вызванных плохим управлением данными, и оценке ценности, добавляемой за счет улучшения управления данными.

Эти изменения не достигаются с помощью технологий (хотя использование программных средств может этому способствовать). Они обеспечиваются за счет тщательного и структурированного подхода к управлению организационными изменениями. Изменения требуются на всех уровнях. Крайне важно, чтобы они управлялись и координировались таким образом, чтобы избежать тупиковых инициатив и потери доверия к функции управления данными и ее руководству.

Культурные изменения требуют планирования, обучения и поощрения. Кроме того, для активизации и вовлечения сотрудников в инициативы, политику и процессы управления данными необходимо создать условия, обеспечивающие осведомленность, распределение ответственности и подотчетность.

Можно выделить десять факторов, играющих ключевую роль в создании эффективной организационной системы управления данными вне зависимости от ее структуры.

1. Наличие куратора в высшем руководстве. Куратор должен понимать инициативу и верить в нее. Он должен быть способен эффективно привлекать других лидеров к изменениям.

2. Четкое понимание. Руководители организации должны обеспечить полное понимание и усвоение того, что такое управление данными и почему это важно для всех заинтересованных сторон, имеющих отношение к управлению данными – как внутренними, так и внешними, – и как это повлияет на их работу.

3. Упреждающее управление изменениями. Применение управления организационными изменениями для создания практик управления данными решает проблемы сотрудников и повышает вероятность того, что создаваемые практики и организационные структуры с течением времени сохранят устойчивость.

4. Согласование позиций руководства. Согласование позиций руководства гарантирует, что существует согласие и единодушная поддержка в отношении необходимости программы управления данными, а также что существует согласие в отношении того, как будет определяться успех. Согласование позиций включает в себя как согласование целей руководства с результатами управления данными и ценностью, так и согласование намерений отдельных лидеров.

5. Коммуникации. Организация должна обеспечить наличие у заинтересованных сторон четкого представления о том, что такое управление данными и почему это важно, а также что меняется и какие потребуются изменения в их поведении.

6. Вовлечение заинтересованных сторон. Отдельные лица и группы, затронутые инициативой по управлению данными, будут по-разному реагировать на новую программу и свою роль в ней. То, каким образом организация взаимодействует с заинтересованными сторонами, как эти стороны с ней общаются, окажет значительное влияние на успех инициативы.

7. Ознакомление и обучение. Образование имеет важное значение для обеспечения управления данными. Различным группам сотрудников (руководители, распорядители данных, владельцы данных, команды ИТ-специалистов) для эффективного выполнения их ролей потребуются различные виды и уровни обучения. Многим людям потребуются тренинги по новым политикам, процессам, методам, процедурам и инструментам.

8. Оценка прогресса освоения. Важно выстроить систему метрик, всесторонне описывающих прогресс внедрения, восприятия и освоения новых принципов и правил управления данными, с целью

объективного контроля выполнения мероприятий, предусмотренных дорожной картой. Показатели, отражающие положительное воздействие управления данными, могут быть связаны с различными аспектами совершенствования дата-центричных процессов, таких как подготовка отчетности по итогам месяца, идентификация рисков, мониторинг эффективности выполнения проектных работ. Инновационный аспект управления данными можно оценивать с помощью показателей повышения качества принятых решений и аналитики за счет повышения качества и достоверности данных.

9. Соблюдение руководящих принципов. Руководящие принципы, такие как принципы управления данными DAMA, служат ориентирами, на основе которых будут приниматься все решения в рамках формируемой организационной системы. Их утверждение – важный первый шаг в создании программы управления данными, обеспечивающей эффективные изменения.

10. Эволюция, а не революция. Во всех аспектах управления данными философский принцип «эволюция, а не революция» помогает свести к минимуму риск при проведении серьезных изменений. Построение организационной системы, которая развивается эволюционно, постепенно совершенствуя методы управления данными в соответствии с целями бизнеса, дает гарантию того, что новые политики и процессы будут приняты, а изменения поведения будут носить устойчивый характер.



ПРАКТИЧЕСКИЙ ПРИМЕР

«Телеком Дубль» предприняла важные шаги по формированию организационных систем руководства и управления данными. С учетом довольно сложной филиальной структуры в качестве ориентира была выбрана гибридная операционная модель.

В компании функционирует управляющий комитет по руководству данными и введена должность директора по данным.

Директор по данным возглавляет совет по руководству данными, управляющий инициативами в области руководства данными (например, разработкой политик или метрик) и разрешением проблемных вопросов.

Определен порядок взаимодействия директора по данным с подразделениями компании по следующим направлениям:

- закупки и контракты;
- бюджет и финансирование;
- нормативно-правовое соответствие;
- разработка информационных систем.

Важно также отметить, что в «Телеком Дубль» для обеспечения всестороннего анализа данных о клиентах на основе их профилей и поддержки взаимодействия с ними создано специальное подразделение «Клиенты 360».

Литература к главе 16

- Aiken P., Gorman M. The Case for the Chief Data Officer: Recasting the C-Suite to Leverage Your Most Valuable Asset. Morgan Kaufmann Publishers Inc., 2013.
- Ladley J. Making Enterprise Information Management (EIM) Work for Business: A Guide to Understanding Information as an Asset. Morgan Kaufmann Publishers Inc., 2010.
- Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.
- Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Д. Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)
- Plotkin D. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance: 2nd Edition. Academic Press, 2020.
- Seiner R. Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success. Technics Publications, 2014.
- Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the

information you need. Entity Group Limited, 2016.

- Soares S. The Chief Data Officer Handbook for Data Governance. MC Press, 2015.

Глава 17. Использование данных и расширение возможностей применения

В главах 10–15 нами были рассмотрены почти все элементы рамочной структуры функций управления данными в привязке к их жизненному циклу (рис. 9.4). Мы сравнили эти элементы с функциональными блоками модели управления цепями поставок (SCOR-модель) и убедились, что схема на рисунке 9.4 вполне отражает модель управления цепочками поставок данных.

Вначале мы подробно остановились на деятельности по осуществлению руководящих и контрольных полномочий в области управления данными – руководстве данными.

Затем был сделан обзор функций, обеспечивающих управление данными на первых двух фазах их жизненного цикла.

1. Планирование и проектирование данных.
2. Обеспечение доступности и обслуживание данных.

Наконец, мы рассмотрели базовые направления деятельности по управлению данными, которые распространяются на весь жизненный цикл информационных активов:

- управление безопасностью данных;
- управление метаданными;
- управление качеством данных.

Все рассмотренные функциональные элементы вносят существенный вклад в цепочку ценности данных и в подготовку информационных активов к ключевой фазе их жизненного цикла – практическому использованию и расширению возможностей применения для достижения целей организации. В этой главе мы остановимся на основных направлениях деятельности в этой фазе, которые играют главную роль в повышении ценности данных и ее извлечении. К ним относятся (см. рис. 9.4):

- использование основных данных;
- бизнес-аналитика;
- наука о данных⁵²⁷.

В предлагаемой Дагласом Лейни упрощенной модели цепочки поставок данных перечисленные направления относятся к завершающему этапу цепочки – «Применению» (см. рис. 7.3).

Заметим, что на рисунке 9.4 к фазе использования данных и расширения возможностей их применения отнесена также функция «Управление документами и контентом». Однако у нас (это уже отмечалось в главе 12) она рассматривается в рамках фазы обеспечения доступности и обслуживания, как это сделано в книге Себастьян-Коулман⁵²⁸.

Нужно также заметить, что здесь мы не обсуждаем монетизацию данных (хотя она выделена на рис. 9.4), поскольку эта тема очень подробно раскрыта в книге Лейни⁵²⁹.

В завершение главы мы поговорим о развитии платформенных решений в области управления данными, играющих важную роль в расширении возможностей их применения.

17.1. Использование основных данных

⁵²⁷ Здесь в рамках деятельности, связанной с наукой о данных, мы также рассматриваем и предиктивную аналитику, которая на рисунке 9.4 выделена как отдельное направление.

⁵²⁸ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

⁵²⁹ Laney D. B. Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage; Routledge; 1st edition, 2017. (Русский перевод: Д. Лейни. Инфономика: информация как актив: монетизация, оценка, управление. – М.: Точка, 2020. – [Библиотека «Айтеко»].)

Вопросы управления основными данными мы довольно подробно рассматривали в главах 12 и 13, здесь лишь еще раз отметим следующее.

Работа с основными данными служит хорошей иллюстрацией того, как использование данных напрямую связано с расширением возможностей их применения. Налаженное управление основными данными позволяет организации лучше понимать особенности объектов и лиц (клиентов, заказчиков, поставщиков, продуктов и т. д.), с которыми она взаимодействует и осуществляет операции.

В процессе своей деятельности организация узнает больше об этих объектах и лицах: что они покупают, что продают, как лучше всего с ними связаться. Новые сведения могут храниться на уровне данных о транзакциях, но организация также должна специально отслеживать и собирать поступающие изменения для поддержания в актуальном состоянии своих основных данных (например, изменения адреса или обновления контактной информации). Транзакционные данные также позволяют им получать дополнительные сведения (например, предпочтения заказчиков или клиентов и истории покупок), которые могут расширить возможности применения основных данных. Взаимосвязи между различными элементами данных и сферами их применения должны учитываться при планировании в любой области управления данными, однако в отношении управления основными данными это предмет особого внимания.

17.2. Бизнес-аналитика

О бизнес-аналитике мы уже немного поговорили в главе 14. Ее инструменты стремительно развиваются и совершенствуются, открывая возможности для перехода от стандартизированной отчетности, диктуемой спецификой используемых информационных технологий, к самостоятельному исследованию данных по направлениям, интересующим бизнес. Можно выделить следующие основные направления развития инструментов BI^{530, 531}.

- **Операционная отчетность** позволяет выявлять и анализировать краткосрочные (помесячные) и среднесрочные (годовые) тенденции и закономерности. Эти инструменты относятся к сфере тактического бизнес-анализа (Tactical BI) и используются для выработки и принятия краткосрочных решений в сфере оперативного управления бизнесом.

- **Управление эффективностью бизнеса** позволяет производить формальную оценку измеримых показателей, соответствующих целям организации. Осуществляется, как правило, на уровне высшего руководства. Эти инструменты относятся к сфере стратегического бизнес-анализа (Strategic BI) и используются для формулировки долгосрочных целей и задач.

- **Приложения для оперативного анализа** предлагают решения для оптимизации различных функциональных областей.

17.2.1. Операционная отчетность

Средства операционной отчетности позволяют генерировать и выводить отчеты непосредственно из транзакционных систем, рабочих приложений или хранилищ данных. Обычно реализуются как функционалы приложений. Очень часто первоначальное применение средств BI – генерирование операционной отчетности, особенно если высокоуровневое распоряжение DW/BI не налажено или в DW содержатся дополнительные по отношению к оперативным или транзакционным данным, учет которых необходим или полезен. Часто операционные отчеты внешне похожи на результаты обработки нестандартных запросов, а на деле представляют собой простые отчеты или вводные для какого-либо рабочего процесса. С точки зрения управления данными ключевым в таких случаях является вопрос о том,

⁵³⁰ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵³¹ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

достаточно ли приложению собственных данных для генерирования отчета, или же ему требуются еще и дополнительные данные из DW или ODS (см. главу 14).

Инструменты исследования данных и формирования отчетности иногда еще называют средствами создания произвольных запросов, поскольку они позволяют пользователям создавать «авторские» отчеты или выборки данных, предназначенных для использования в качестве вводных другими пользователями или процессами. Строгого соблюдения каких-либо стандартных требований к структуре или формату документа в данном случае не предъявляется, поскольку речь идет не о счетах-фактурах или чем-то подобном. Зато пользователям часто интуитивно хочется включать в такие отчеты графики и таблицы. Зачастую созданные бизнес-пользователями с помощью произвольных запросов отчеты оказываются настолько удачными, что утверждаются в качестве стандартной формы внутриорганизационной отчетности по затрагиваемому в них кругу вопросов.

Требующиеся бизнесу операционные отчеты часто не совпадают с отчетами, генерируемыми по стандартным запросам, которые обычно используют в качестве источника DW или как предназначенную для соответствующего бизнес-подразделения витрину данных. Кроме того, стандартные отчеты обычно разрабатываются ИТ-специалистами, а произвольные – продвинутыми бизнес-пользователями с помощью программных средств построения запросов. При необходимости созданные пользователями запросы и отчеты можно утверждать к регулярному использованию в рамках отдела или всего предприятия.

Производственная отчетность часто выходит за рамки DW или BI и включает запросы к транзакционным системам с целью получения вводных для таких оперативных документов, как счета-фактуры или банковские выписки. Запросы и форматы производственных отчетов обычно разрабатываются ИТ-специалистами.

Традиционные инструменты BI включают ряд стандартных средств наглядного представления данных в виде таблиц, секторных и линейных графиков, столбцов или гистограмм. Помимо статичных форматов визуализации, используемых в отчетах для публикации, возможны также динамические и даже интерактивные форматы в онлайн-отчетах, вплоть до поддерживающих масштабирование, навигацию по уровням детализации и применение фильтров с целью упрощения анализа данных в визуализированном представлении. Может быть предусмотрено также и пользовательское переключение между различными типами графиков или режимами их отображения.

17.2.2 Управление эффективностью бизнеса

Управление эффективностью бизнеса (Business Performance Management, BPM) – это набор интегрированных процессов организации и приложений, разработанных для оптимизации исполнения бизнес-стратегии. Стандартный набор поддерживает формирование бюджета, планирование, бухгалтерский учет и сводную финансовую отчетность. Нарботки в этом сегменте имеются огромные, поскольку производители программного обеспечения как для управления предприятием (ERP), так и для BI видят в данной области огромные резервы роста, к тому же грань между бизнес-аналитикой и управлением эффективностью все более стирается. Частота приобретения клиентами решений в области BI и управления эффективностью от одного и того же разработчика зависит от возможностей поставляемых им продуктов.

Технология BPM позволяет приводить процессы в соответствие с организационными целями. Ключевые элементы BPM – измерения и петля положительной обратной связи. В сфере BI это приняло форму множества приложений для различных стратегических областей деятельности предприятия – бюджетного планирования, прогнозирования и планирования ресурсов. Другая специализация BI сформировалась строго внутри этой области и включает создание карт балльной оценки в связке с приборными панелями для интерактивного информирования пользователей. Как и в автомобиле, на приборную панель, находящуюся в поле зрения конечного пользователя, выводится сводка текущих значений важнейших показателей.

17.2.3. Приложения для оперативного анализа

Аналитические приложения работают по принципу извлечения данных из хорошо известных систем, таких как стандартные ERP или модели данных для представления в витринах, и переработки их в предустановленные показатели и форматы для вывода в отчеты или на информационные панели. По сути,

бизнесу предлагаются готовые решения для оптимизации различных функциональных областей (например, управления персоналом) или встраивания в отраслевую вертикаль (например, аналитика розничного рынка). Приложения различных типов могут включать функции анализа клиентов, финансов, цепочек поставок, организации производства или управления персоналом.

Как отмечалось в главе 14, среди приложений для оперативного анализа особо выделяются инструменты онлайн-аналитической обработки (online analytical processing, OLAP), обеспечивающие высокопроизводительную обработку многомерных аналитических запросов. Термин OLAP возник отчасти в противовес термину OLTP, используемому для обозначения онлайн-обработки транзакций. Обычно выдача данных в ответ на запросы OLAP происходит в матричном формате. Измерения определяются столбцами и строками матрицы, на пересечении которых выводятся факторы или значения. Концептуально это представление иллюстрируется как куб данных. Многомерный анализ с кубами особенно полезен там, где у аналитиков имеется хорошее представление об общей картине и структуре данных, а разобраться хочется с динамикой и сводной статистикой.

Традиционная область применения OLAP – финансовый анализ, ведь специалисты в этой области привыкли иметь дело со сводными таблицами данных, упорядоченными в рамках хорошо известных иерархий, выискивая и анализируя тенденции и закономерности; а кубы данных позволяют с легкостью переходить на иную шкалу измерений или масштабов даты и времени (годовые, квартальные, месячные, недельные, суточные, почасовые показатели), организационной структуры (мир, регион, страна, отрасль, компания, подразделение) или иерархии продуктов (категория, линия, наименование продукта). Многие пакеты программного обеспечения для BI сегодня используют OLAP-кубы в качестве одной из базовых моделей.

Краткие сведения об основных архитектурах OLAP-систем и принципах организации портфеля BI-приложений приведены в главе 14.

17.3. Наука о данных

В главе 14 был описан процесс осуществления деятельности в области науки о данных. Здесь мы рассмотрим наиболее важные из применяемых учеными по данным методов:

- машинное обучение;
- анализ настроений;
- интеллектуальный анализ данных и текстов;
- предиктивная аналитика;
- предписывающая аналитика;
- методы анализа неструктурированных данных;
- визуализация данных;
- объединение данных.

17.3.1. Машинное обучение

Машинное обучение исследует методы построения алгоритмов, реализованных в программном обеспечении. Можно рассматривать машинное обучение как синтез методов неконтролируемого самообучения (часто называемых извлечением информации – data mining) и методов контролируемого, или управляемого, обучения, которые имеют глубокие математические корни, в том числе из статистики, комбинаторики и оптимизации систем. Начала формироваться и третья ветвь – так называемое обучение с подкреплением без учителя: задаются целевые параметры, и система упражняется в их соблюдении (пример: автопилот транспортного средства). Программирование машин на быстрое усвоение повторяющихся структур запросов и адаптацию к изменениям наборов данных привело к появлению одноименного раздела «машинное обучение» и в области больших данных, где эта концепция получила новое применение. Процессы прогоняются, результаты сохраняются, а затем используются при последующих прогонах для уточненной настройки процесса, и такие итерации повторяются до получения результата желаемого уровня точности и детализации.

Машинное обучение занимается структурным построением алгоритмов познания и усвоения знаний. Выделяют три типа таких алгоритмов.

1. **Обучение с учителем** основано на применении обобщенных правил (пример: настраиваемый фильтр спама в почтовом приложении).
2. **Обучение без учителя** основано на выявлении скрытых паттернов, связей, закономерностей (т. е. собственно интеллектуальный анализ данных).
3. **Обучение с подкреплением** основано на достижении цели (например, выигрыша шахматной партии).

Статистическое моделирование и машинное обучение используют также для автоматизации нереализуемых или слишком затратных процессов в рамках исследовательских и проектных работ, когда требуется, например, методом проб и ошибок подобрать ключ к огромному набору данных, повторяя цикл экспериментальной обработки, анализа результатов и исправления ошибок. Такой подход позволяет значительно ускорить получение ответа, что и стимулирует организации к инициативам по поиску глубинных закономерностей посредством многократного повторения затратно эффективных процессов.

Хотя машинное обучение и использует весьма новые способы получения данных, в этой новой области знания должны соблюдаться все традиционные принципы этичного обращения с данными, в частности и прежде всего – принцип прозрачности. Появились научные доказательства того, что метод обучения нейронных сетей с помощью глубокого погружения работает. Они учатся и постигают мир. Однако не всегда ясны механизмы их обучаемости. Чем сложнее алгоритмы, лежащие в основе этих процессов, тем менее они прозрачны – и начинают функционировать в режиме «черного ящика». Чем больше переменных учитывают самообучаемые нейронные сети и чем более абстрактными делаются сами эти переменные, тем больше реализуемые ими алгоритмы испытывают пределы возможностей человека понимать и интерпретировать логику машинного мышления. Необходимость обеспечения прозрачности принятия решений по мере дальнейшего совершенствования функциональности неконтролируемого самообучения и его применения во все более широком спектре ситуаций, вероятно, будет только возрастать^{532, 533}.

17.3.2. Анализ настроений

Мониторинг медиа и анализ текста относятся к автоматизированным методам извлечения аналитической информации из больших массивов неструктурированных и слабо структурированных данных, включая страницы отзывов, соцмедиа, блоги и новостные веб-сайты. Делается это для того, чтобы понять и обобщить мнения людей и выявить преобладающее в различных социальных группах отношение к брендам, продуктам или услугам, а также любым другим темам или явлениям. Используя алгоритмы обработки естественного языка, синтаксического и лексического разбора предложений или формулировок, средства семантического анализа позволяют выявлять не только доминирующую в высказываниях эмоциональную окраску, но и динамику ее изменения во времени, что открывает возможность предсказывать вероятные сценарии дальнейшего развития событий.

Проиллюстрируем этот подход на простейшем примере поиска и подсчета статистики частоты употребления ключевых слов в опубликованных отзывах о продукте. Если в комментарии присутствуют слова «отличный», «восторг» или «замечательно», вероятно, это позитивный отклик, а присутствие слов «плохой», «дрянь», «гадость» может служить признаком негативного отношения. Распределив отзывы по категориям, можно выяснить преобладающее в целевом сообществе (например, в данной соцсети или блоге) отношение. Но реальные чувства и эмоции, вызываемые предметом обсуждения, не так легко бывает уловить по причине того, что любое ключевое слово, будучи вырванным из контекста, может быть интерпретировано неверно. Например, слово «ужасно» вроде бы указывает на негативное отношение к ресторану, а в отзыве написано: «Ужасно вкусно!» А формально позитивную характеристику «сказочно»

⁵³² DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵³³ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

можно найти в возмущенном отзыве: «Сказочно нерасторопное обслуживание!» Поэтому семантический анализ эмоциональной окраски должен интерпретировать слова только в контексте. А это уже требует понимания смыслового значения отзыва или комментария. Для правильной интерпретации смысла написанного часто требуются функции обработки естественного языка, реализованные на сегодняшний день лишь в суперсистемах уровня IBM Watson^{534, 535}.

17.3.3. Интеллектуальный анализ данных и текстов

Интеллектуальным анализом данных (или извлечением информации, data mining) принято называть применение к массивам разнородных данных разнообразных алгоритмов выявления скрытых структурных закономерностей. Интеллектуальный анализ данных постепенно отделился от машинного обучения и сделался отдельной подобластью исследований по созданию искусственного интеллекта. Теория интеллектуального анализа данных формально относится к методологии статистического анализа, известной под названием «обучение без учителя», которая предусматривает применение к набору данных неких алгоритмов изучения, никак не связанных с ожидаемым или желаемым результатом. В то время как стандартные средства генерации запросов и отчетов формулируют вполне конкретные требования к данным, средства интеллектуального анализа данных помогают раскрывать неизвестные ранее взаимосвязи через выявление повторяющихся структур (паттернов). Извлечение данных – ключевое направление работ на этапе первичного изыскания возможностей, поскольку позволяет оперативно идентифицировать поддающиеся изучению элементы обрабатываемого массива данных, выявлять ранее неизвестные и уточнять нечеткие или неклассифицированные связи, закладывая структурную основу классификации элементов изучаемых данных.

В сочетании с семантическим и структурно-лингвистическим анализом текстовой информации интеллектуальный анализ данных позволяет автоматически классифицировать данные по признакам их содержания и интегрировать полученные классификации в онтологии, составляемые по мере накопления данных под общим руководством экспертов в предметной области. Таким образом, появляется возможность анализа электронных текстов в различных средах и форматах без их реструктурирования или конвертирования. Накапливаемые онтологии можно подключать к информационно-поисковым системам, что даст пользователям и приложениям возможность получать доступ к этим документам через поисковые запросы.

Извлечение данных и интеллектуальный анализ текстов основаны на использовании ряда стандартных технических приемов, включая описанные ниже.

Профилирование заключается в описании характерных типов поведения людей, групп или организаций и используется для определения признаков нормального поведения с целью выявления серьезных отклонений от нормы, например в приложениях по отслеживанию мошеннических операций или попыток проникновения в системы. Результаты профилирования служат входными данными для многих компонентов, работающих по принципу самообучения.

Сокращение избыточных данных позволяет заменять исходные, излишне детализированные наборы данных обобщенными, где сохраняются лишь ключевые характеристики или категории, что заметно ускоряет и упрощает обработку и анализ.

Ассоциирование часто встречающихся в связке друг с другом элементов – еще один стандартный алгоритм выявления взаимосвязей, применяемый в интеллектуальном анализе данных. Ассоциативные связи могут использоваться, например, для накопления статистики часто встречающихся наборов элементов, выявления скрытых правил, анализа конъюнктуры локальных рынков. А рекомендательные системы в интернете без использования подобных алгоритмов не обходятся.

Кластеризация: группировка элементов в кластеры по признаку близкого сродства или общности

⁵³⁴ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵³⁵ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

неких характеристик упрощает и ускоряет статистический анализ типичных схем и стереотипов поведения. Классический пример кластеризации – сегментация потребительского рынка.

Самоорганизующиеся карты – метод кластерного анализа нейронных сетей, известный также под названием самоорганизующихся карт Кохонена или топологически упорядоченных карт. Их использование позволяет снизить размерность пространства оценки без ущерба для результатов аппроксимации. Устранение избыточных пространственных измерений, отметим, по эффективности не уступает изъятию вырожденных переменных из алгебраических уравнений – и решать проще, и результат нагляднее^{536, 537}.

17.3.4. Предиктивная аналитика

Предиктивной аналитикой называют подраздел обучения с учителем, в рамках которого пользователи пытаются смоделировать элементы данных и предсказать будущие исходы по оцениваемым вероятностям событий. В методах теории вероятностей и математической статистики прогнозная аналитика, однако, имеет много общего с обучением без учителя в части прописывания, например, предельно допустимых отклонений полученных результатов от предполагаемых, после чего требуется пересмотр гипотез.

Таким образом, предиктивная аналитика основана на использовании обычных вероятностных (стохастических) моделей обработки вводных данных (включая исторические) для определения вероятности будущих событий (покупок, ценовых изменений). При получении информации, выходящей за рамки текущей модели, сама же модель и запрашивает у организации порядок дальнейших действий. Фактором запуска может служить любое событие: заказ в интернет-магазине, текст в новостной ленте, образ в системе распознавания лиц, непредвиденный всплеск спроса на услуги. Пусковым моментом могут являться и внешние факторы. Например, появление негативных материалов о компании в СМИ – верный признак скорого снижения биржевых котировок ее акций. А способность прогнозировать динамику биржевых котировок по новостям – отличное функциональное свойство средств аналитики данных с точки зрения игроков на фондовых рынках.

Зачастую превышение критического порога потока каких-либо характерных данных в режиме реального времени (например, биржевых сделок или обращений в экстренную службу) служит причиной для запуска цепи всевозможных последствий в динамично меняющейся и нестабильной среде. Мониторинг потока событийных данных позволяет устанавливать пороги счетчиков критических событий, определяемых в рамках модели и служащих сигналом для выдачи предупреждения или запуска каких-либо действий.

Запас времени, которое остается в распоряжении у получателей сигнала о прогнозируемом событии до фактического наступления этого события, нередко бывает мизерным (вплоть до долей секунды). Поэтому инвестиции в технологии быстрого реагирования (в частности, резидентные базы данных, широкополосные каналы связи и даже физический перенос ЦОД в непосредственную близость к объекту – источнику данных) оправдываются, если позволяют реально повысить способность к прогнозированию и оперативному реагированию на прогноз.

Простейшая модель прогнозирования – статистическая. Существует множество методик статистического прогнозирования, основанных на выявлении тенденций с экстраполяцией или регрессионном анализе, но в любом случае требуется сглаживание. Простейший вариант сглаживания данных реализуется путем расчета скользящего среднего или средневзвешенного значения. В специфических случаях могут применяться более сложные техники сглаживания, такие как расчет экспоненциального скользящего среднего, что позволяет управлять коэффициентом сглаживания (фильтрации флуктуаций). Для начала можно применить один из методов регрессионного анализа – метод наименьших квадратов, но в любом случае требуется несколько пробных прогонов для подбора оптимального коэффициента сглаживания. Существуют модели с двумя и более фильтрами

⁵³⁶ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵³⁷ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

экспоненциального сглаживания, позволяющие учитывать, например, недельные колебания на фоне сезонных^{538, 539}.

17.3.5. Предписывающая аналитика

Предписывающим анализом называют прогнозный анализ, дополненный определениями корректирующих воздействий на ситуацию с целью изменения конечных результатов, а не ограничивающийся простым их прогнозированием. Таким образом, предписывающая аналитика позволяет предсказывать, что случится, когда это случится и по совокупности каких факторов это случится. Будучи способным демонстрировать последствия различных сочетаний решений, предписывающий анализ позволяет моделировать их комбинации с целью максимизации выигрыша или минимизации риска. Методы предписывающего анализа удобны тем, что предусматривают возможность непрерывной подачи на вход скорректированных вводных и перерасчета прогнозов с выдачей скорректированных предписаний. Это повышает и точность прогноза, и результативность предписаний^{540, 541}.

17.3.6. Методы анализа неструктурированных данных

Анализ неструктурированных данных основан на сочетании различных методов анализа текстов, ассоциаций, кластеров и прочих вышеописанных методов обучения без учителя, помогающих кодифицировать большие наборы слабо структурированных данных. Могут использоваться и методы обучения с учителем – например, чтобы задать направление, ориентацию и наставления машинному мышлению на правильный подход к кодированию выявляемых структурных зависимостей – и часто лишь человеческое вмешательство позволяет избежать невнятности формулировок или разрешить неоднозначности.

Значение анализа неструктурированных данных возрастает пропорционально нарастанию их доли в мировом информационном пространстве. Бывает, что анализ какого-либо явления просто невозможен без включения в аналитическую модель неструктурированных данных. Однако анализ неструктурированных данных осложняется необходимостью предварительного отделения интересующих исследователей данных от лишних элементов.

Сканирование и тегирование – единственный способ выуживания полезных неструктурированных данных из озера, позволяющий отфильтровать их от воды и привязать к структурированным данным. Тем не менее тут возникает следующая проблема: какими тегами маркировать данные, не зная заранее их содержания, и как определить условия тегирования? Ответ может быть получен только итерационным путем: по мере выявления реальных условий тегирования уточняются и начинают присваиваться теги, а по мере поглощения и освоения тегированных данных аналитики проверяют правильность условий тегирования, анализируют выловленные данные – и постепенно уточняются и согласуются все условия

⁵³⁸ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵³⁹ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

⁵⁴⁰ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁴¹ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

тегирования и структура тегов, а по мере надобности могут добавляться и новые теги^{542, 543}.

17.3.7. Визуализация данных

Визуализация данных – процесс интерпретации концепций, идей и фактов через наглядные представления, включая фотографии, рисунки, коллажи и всевозможные графики, а также схемы. Визуализация упрощает понимание иллюстрируемых данных, обеспечивая наглядность и лаконичность их сводного (например, графического) представления. Визуализация позволяет предельно сжато и доходчиво отображать наиболее характерные данные с целью навести зрителей на полезные выводы о скрытых возможностях, рисках или смыслах.

Визуальные представления могут быть как статичными (например, в формате иллюстрированного отчета), так и анимированными, динамично обновляемыми и даже интерактивными, т. е. позволяющими конечному пользователю переходить на различные уровни детализации, накладывать фильтры и иным образом упрощать себе визуальный анализ данных. В качестве варианта может предусматриваться и переключение пользователем режима отображения данных в инновационные форматы, такие как интерактивные географические карты и динамические ландшафтные пейзажи данных.

Анализ данных уже давно немыслим без средств визуализации. Все традиционные инструменты бизнес-анализа обязательно включают широкий выбор средств визуального представления данных: таблицы, всевозможные линейные и круговые, плоскостные и объемные, столбчатые и полосчатые графики, гистограммы. С ростом спроса на наглядные данные безостановочно совершенствуются средства их визуализации.

По мере роста зрелости информационной аналитики новые способы визуального отображения данных становятся важным стратегическим преимуществом. Новый взгляд на данные позволяет выявить новые связи и закономерности, а следовательно, и новые возможности для бизнеса. По мере дальнейшего развития и совершенствования средств визуализации организациям придется взращивать такие команды бизнес-аналитиков, которые смогут обеспечивать им конкурентоспособность во все более компьютерно-управляемом в потоковом режиме мире. И вот тогда бизнес-аналитическими отделами будут крайне востребованы эксперты с навыками визуализации – знатоки данных, художники данных, визионеры данных – в дополнение к традиционно ценящимся архитекторам и разработчикам моделей данных. Это будет более чем оправданно, если помнить о рисках, проистекающих от искажающих восприятие обманчивых визуальных представлений^{544, 545}.

17.3.8. Объединение данных

Средства получения данных из различных источников и служб позволяют создавать различные агрегированные представления данных для нужд визуализации или анализа. Многие инструменты виртуализации поддерживают агрегирование через функциональность связывания данных из различных источников объединяющими элементами, т. е., по сути, тем же приемом, который традиционно использовался в реляционных моделях для связывания, к примеру, объекта и описания через внешний ключ. Техническая возможность создания различных данных весьма полезна для получения

⁵⁴² DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁴³ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

⁵⁴⁴ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁴⁵ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

пользовательских представлений и идеально подходит для реализации задач, которые возникают на фазах раскрытия источников или разведки ресурсов данных, позволяя получать быстрые и наглядные результаты. Этот метод может быть применен в веб-приложении, поскольку позволяет организовывать обмен защищенными нарезками, содержащими персональные или конфиденциальные данные, между поставщиками или провайдерами информационных услуг. В сочетании с алгоритмами обучения искусственного интеллекта такие агрегированные представления помогают выявлять интернет-сервисы, оснащенные интерфейсами с поддержкой обработки естественного языка^{546, 547}.

17.4. Развитие платформенных решений в области управления данными

Если продолжать разговор о расширении возможностей применения данных, то следует сказать о том, что в последние несколько лет наблюдается следующая тенденция: технологические компании переходят от отдельных решений и сервисов к платформенным решениям, т. е. решениям построения технологических экосистем. Это позволяет существенно расширить функциональность своих сервисов и быстро вводить новые услуги для своих абонентов/подписчиков, а также обеспечить сторонним решениям площадку для размещения и доступа сервисов для своих абонентов.

Этот эволюционный путь, основу которому положили такие гиганты, как Apple и Google, развивался все последние годы. Теперь же, с появлением COVID-19, многие технологические компании приняли такой путь как существенное преимущество для своих продуктов и сервисов. Все встали перед сложным выбором: либо оставаться в своей нише и опираться на внешние платформенные решения, что существенно ограничивает возможности развития бизнеса и требует серьезных усилий для борьбы с возрастающей конкуренцией, либо развивать свои экосистемы и платформенные решения, что позволяет быстрее расти, но требует серьезных инвестиций. Помимо этого, при выборе второго подхода очень велик риск того, что без технологичной основы создание цифровых сервисов может стать очень дорогим и даже разорительным.

Речь идет о программном обеспечении, которое предоставляет большое количество услуг и функциональных возможностей в рамках одного продукта, а зачастую предлагает платформу для включения сервисов для внешних разработчиков. Самые известные платформы это Apple iOS, Google Apps для пользовательского сегмента (B2C); SAP, ORACLE, «Юнидата» для корпоративного (B2B). Создание платформенных решений для отдельных продуктов становится существенным конкурентным преимуществом, позволяющим быстро выводить на рынок новые продукты и сервисы, обеспечивать эффективную интеграцию между отдельными решениями и оптимизировать единые требования к стеку используемых технологий.

Платформенные решения также позволяют пользоваться своей богатой функциональностью и уже реализованными сервисами как собственным разработчикам, так и внешним компаниям и специалистам. Одним из значимых преимуществ для новых сервисов при размещении его на платформе является мгновенный доступ к клиентам, уже использующим решение, либо в режиме реального времени, либо при выходе очередного релиза (версии ПО).

Важнейшее преимущество использования платформенных технологий – возможность работать с единым массивом данных, формирующихся во всех функциональных блоках, подсистемах и приложениях. Архитектура сложных платформ может быть сформирована из нескольких функциональных решений и отдельных платформенных решений. Например, решения класса маркет-приложений, помимо той ее части, которая доступна конечному пользователю, включает в себя платежные платформы, платформы разработчиков, платформу управления данными и другие. Стоит отметить, что управление данными во многих решениях становится ключевым конкурентным преимуществом. Более того, введение подобных

⁵⁴⁶ DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017. (Русский перевод: DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Dama International. – М.: Олимп-Бизнес, 2020.)

⁵⁴⁷ Sebastian-Coleman L. Navigating the Labyrinth: An Executive Guide to Data Management, First Edition. Technics Publications, 2018.

систем – необходимый критерий существования любой компании на рынке. И роль системы управления данными будет возрастать.

Объем обрабатываемых данных в некоторых платформах только в части критически важных записей (основных данных) уже составляет миллиарды объектов, а объем связанных с ними записей – сотни миллиардов единиц. При этом транзакционных данных обрабатывается на порядки больше, и их привязка к основным данным осуществляется в режиме реального времени (онлайн). Подобный беспрецедентный объем данных предъявляет уникальные технологические, архитектурные, функциональные требования к платформам управления данными, в первую очередь в части полноты, качества и инструментария работы с основными данными, являющимися по сути скелетом работы с данными.

Консолидация данных в подобных объемах позволяет получить значительный синергетический эффект, обеспечивающий большое количество инновационных сервисов, удобство пользовательских интерфейсов, формирование различных цифровых профилей, оптимальную архитектуру решений, технологическую и экономическую эффективность ее модернизации и развития, повышение качества интеграционных процессов и другие преимущества. Вместе с тем широкое распространение платформенных решений уже запустило процесс глобального изменения бизнес-процессов целых отраслей, а в будущем кардинально изменит также и экономическую модель работы многих национальных и международных компаний, что приведет к формированию совершенно новой структуры мировой экономики. Об этом мы более подробно писали в предыдущем разделе.

В качестве очень красноречивого и близкого каждому пользователю мобильных устройств примера рассмотрим возможный сценарий развития магазина приложений, используемых в смартфонах (речь идет об App Store в случае смартфонов под управлением операционной системы iOS и Google Play в случае смартфонов на Android). Компании Apple и Google весьма четко определили модель функционирования своих решений и возможности получения доходов не только от компаний – разработчиков сторонних приложений, оплачивающих размещение своих продуктов, сервисов в магазине приложений и впоследствии работающих в смартфонах в операционных системах, но и с моделью монетизации собственных сервисов. В основу создания собственных сервисов Apple и Google был положен принцип «близости к абоненту», ведь тот кто ближе к пользователю, тот и управляет его потребностями и желаниями.

Более 20 лет назад эта модель бизнеса позволила операторам мобильной связи быстро вытеснить с рынка предоставления услуг связи операторов фиксированной связи, вынудив их активно развивать новые сегменты услуг, в первую очередь телекоммуникационных, используя и модернизируя свои сети, что привело фактически к переходу операторов связи в сегмент предоставления услуг высокоскоростного доступа в интернет для домашних хозяйств и офисов. В то же время многие люди практически полностью перестали пользоваться домашним телефоном, окончательно переведя свое общение на мобильный телефон.

Мобильные операторы связи в конце 1990-х – начале 2000-х годов с учетом технологических ограничений, могли предложить исключительно голосовую связь и короткие СМС-сообщения. По мере развития телекоммуникационных технологий и увеличения скорости передачи данных сети мобильной связи существенно повысили свои возможности, что отразилось в первую очередь на расширении функциональных возможностей абонентских устройств, и постепенно только появившиеся смартфоны вытеснили телефоны сотовой связи. В дальнейшем развитие телекоммуникационных технологий позволило абонентам мобильной связи пользоваться все большим количеством сервисов, которые были доступны посредством версий интернет-браузеров для мобильных устройств. Это время как раз и стало периодом зарождения новой технологической концепции, именно тогда заговорили о том, что телекоммуникационные сети – это фактически трубы для высокоскоростной передачи данных, а производители абонентских устройств становятся гораздо ближе к абоненту.

Именно тогда крупнейшие технологические компании мира (разработчики программного обеспечения, производители смартфонов, телекоммуникационные компании) приступили к активной разработке своих операционных систем для смартфонов, поскольку телекоммуникационные сети уже стали «трубой», а смартфоны ожидало будущее просто «аппаратной абонентской части этой трубы». Сложно оценить объем инвестиций, направленных на создание этих операционных систем, но порядок выделенных средств, вне всяких сомнений, измерялся десятками миллиардов долларов. Столь существенные

инвестиции объясняются огромными затратами на всех стадиях исследований и разработок, а также крайне высокой важностью задачи, поскольку для многих компаний речь шла фактически о жизни и смерти или о дальнейшем динамичном развитии на протяжении многих лет, либо стагнации на долгие годы.

В этой крайне дорогой гонке технологий победу одержали компании Apple и Google с той разницей, что iOS установлена исключительно на продукции Apple, а Android используется большинством производителей смартфонов. Конечно, и другие крупные технологические компании не оставили попыток создать свои операционные системы для мобильных устройств, и, возможно, в ближайшем будущем мы станем свидетелями широкого распространения решений, разработанных компаниями Microsoft, Huawei, Samsung, NTT Docomo и других технологических лидеров, но в настоящее время разрыв с двумя лидерами рынка огромен.

Чем же обусловлено такое внимание крупнейших компаний к этой области? Дело в том, что, как мы уже говорили, тенденция, обусловленная принципом «тот, кто ближе к абоненту/пользователю, тот и управляет его потребностями/желаниями», была основным подходом в индустрии, а учитывая темпы развития цифровых технологий и увеличение скоростей передачи данных в сетях мобильной связи, многие люди стали использовать свои смартфоны в качестве инструментария для получения многих онлайн-сервисов, которые ранее были доступны исключительно в офлайн-формате. Сейчас смартфон для многих уже стал платежным средством, навигатором, переводчиком, устройством доступа к социальной сети, агрегатором новостей, инструментом для работы, клиентом почтовых сервисов, средой для получения государственных цифровых сервисов, средством покупки в интернет-магазинах и продвинутым средством общения, а зачастую и игровым устройством. И функциональные возможности благодаря выходу все новых и новых приложений продолжают расширяться. При этом компании продолжают вкладывать значительные инвестиции в развитие своих платформ, расширяя не только технологические и функциональные возможности, но и свое географическое присутствие.

В настоящее время Apple и Google накопили столь значительный опыт и объем данных о своих клиентах и разработчиках, что анализ этих знаний позволяет им оперативно корректировать стратегию развития решений. Эти уникальные знания предоставляют им возможность точно планировать свои огромные инвестиции в самые востребованные и экономически выгодные сервисы, зачастую подменяя внешние приложения на сервисы, встраиваемые в платформу, такие как приложения, способные собирать различные данные об активности пользователей, медиаприложения, новостные агрегаторы, платежные системы, электронные книги, навигационные сервисы, почтовые клиенты и многие другие. В будущем накопленный объем данных о пользователях, а также возможности инвестирования огромных средств приведет к появлению в платформенных решениях Apple и Google большинства самых востребованных сервисов в виде встроенных возможностей. В самое ближайшее время пользователи мобильных устройств Apple и операционной системы Google смогут отказаться от продуктов онлайн-банков в пользу платежных и банковских приложений Apple и Google. Также два технологических лидера смогут предложить пользователям сервисы такси, электронных магазинов, медиабibliothек, сервисы бытовых услуг, сферы туризма и гостеприимства, транспорта, аренды различных транспортных средств, покупки и аренды недвижимости, сферы питания и сервисов доставки.

Речь идет о совершенно новом качестве предоставляемых услуг, не имеющих географических границ, совершенно ином уровне удобства пользователей. Все это позволит, с учетом агрегации огромных денежных средств, предоставить уникальную программу лояльности, объединяя все сервисы для каждого конкретного цифрового пользователя, собирая еще больше данных о своих клиентах, обогащая цифровой профиль информацией о предпочтениях пользователей, профиле их потребления, поездках, и многих других аспектах жизни клиентов. Все эти изменения в экономических моделях большинства пользовательских сервисов и переход их к цифровому формату и глобализации приведет к кардинальным изменениям деятельности большинства компаний. Компании, способные уже сейчас перестраивать свою работу, окажутся в более выгодном положении. Одновременно многие компании будут вынуждены прекратить свою деятельность.

В ближайшее время нас также ждет увеличение разрыва между национальными компаниями и глобальными корпорациями практически во всех сферах деятельности: в масштабах исследований и разработок, развития технологических решений и платформ, сфере поглощений и слияний, проведении маркетинговых исследований и программ, выведении на рынок новых продуктов и их развитии,

управлении стоимостью решений.

Параллельно стоимость развития высокотехнологических решений будет расти. Однако стоимость создания новых решений на базе платформ существенно ниже обособленных решений, создаваемых фактически с нуля и не опирающихся на ранее созданные технологии платформы. К тому же платформы используют огромный синергетический эффект от консолидации данных различных технологических и функциональных подсистем. Учитывая все это, можно констатировать, что будущие глобальные технологические лидеры – это компании, строящие и развивающие платформенные решения во всех направлениях высоких технологий. В ближайшее время национальные и локальные технологические решения и платформенные системы с большей долей вероятности либо станут неконкурентоспособными и исчезнут, либо станут составной частью глобальных компаний, за исключением уникальных решений, формирующих новые технологические сферы и рынки потребления.

На примере Apple и Google мы уже рассмотрели тенденции, заметные в сегменте продуктов/сервисов для потребителей (B2C). Похожие изменения происходят в сегменте бизнеса для бизнеса (B2B). Для корпоративного сегмента платформы больше напоминают конструктор Lego, где заказчик имеет возможность использовать различные функциональные подсистемы по мере необходимости, расширяя их масштаб, учитывая быстродействие и дополняя новыми подсистемами. В решениях для B2B платформенный слой – важнейшая составляющая, поскольку обеспечивает все подсистемы универсальной и общей для всей системы интеграционной составляющей, инструментарием для работы с данными, базовой аналитикой и другой общесистемной функциональностью. При этом в корпоративных решениях вопросы интеграции как на уровне взаимодействия функциональных подсистем, так и на уровне данных критически важны, они позволяют не только корректно работать функциональным подсистемам, но и обеспечивать качественную работу всей платформы как единого механизма с единым массивом качественных, полных и непротиворечивых данных, едиными для всей платформы бизнес-процессами, единым инструментарием настройки отдельных подсистем и системы в целом.

Еще один основополагающий слой работы современных платформенных решений – инструментарий работы с данными. Более того, в самом ближайшем будущем это, видимо, и станет фундаментом всех без исключения платформ. Именно эта основа позволяет функционировать всей системе, поскольку любая функциональность, даже самая продвинутая, без данных не актуальна, да и любые ошибки в данных в корпоративном сегменте могут привести к катастрофическим последствиям. Во многом по этим причинам роль решений по работе с данными с каждым годом возрастает, и если раньше эту предметную область воспринимали как составную часть крупных ИТ-решений, то сейчас это уже отдельная область знаний, далеко вышедшая за рамки прежних стандартов. Так, большинство глобальных и национальных высокотехнологических компаний, работающих в области корпоративного программного обеспечения, уже предлагают заказчикам решения в области управления данными, среди них такие компании, как IBM, SAP, Oracle, Informatica, «Юнидата» и другие.

По мнению многих аналитиков, в будущем навыки работы с инструментарием по управлению данными станут обязательным требованием большинства работодателей не только для ИТ-специалистов, но и для людей других профессий, поскольку в современном мире практически все так или иначе состоит из данных. В этой области, как и в сегменте B2C, но с небольшим отставанием, заметны те же тенденции усиления конкурентной борьбы и формирование компаний/решений – глобальных лидеров. От платформенных решений, функционирующих на уровне предприятий и корпораций, уже сейчас по большей части зависит эффективность этих организаций, а работоспособность этих систем – один из ключевых критических условий деятельности, поскольку любые существенные неисправности могут привести к огромным потерям, измеряемым как денежными средствами, так и потерей доли рынка, уроном имиджу и другим негативными последствиями, способными привести к краху всего бизнеса.

Сегодня подавляющее большинство крупных международных компаний либо уже работает на основе платформенных решений различной функциональной направленности, либо рассматривает возможность их внедрения. Одной из интересных особенностей последнего времени становится развитие систем, позволяющих размещать в решения вендора отдельные функциональные подсистемы, созданные сторонними компаниями – разработчиками программного и аппаратного обеспечения. Фактически технологические платформы для корпоративного сегмента (B2B) превратятся в решение, составной частью которого будет маркетплейс, аналогичный решениям App Store и Google Play. В дополнение к знакомой

нам функциональности отдельных платформ для корпоративного сегмента постепенно приходят решения интернет вещей (Internet of things – IoT), позволяющих собирать огромные массивы данных с различных устройств и датчиков, а также работать с этими данными и устройствами. Подобная функциональность уже служит существенным конкурентным преимуществом в сегменте как B2B, так и B2C. Стоит отметить, что в ближайшем будущем большинство платформенных решений для корпоративного сектора станут облачными и получат еще большее распространение. Уже сейчас сложно представить себе крупную компанию, не использующую решения класса ERP, CRM, MDM и других. Учитывая роль и объем данных, накопленных и постоянно пополняющихся в этих системах, можно предположить, что одним из основных направлений дальнейшего развития информационных систем, существенно повышающих эффективность бизнеса, станут многофункциональные платформы управления данными, в которых на единой платформе будут собраны различные подсистемы для работы с данными.

По мере развития аппаратного обеспечения увеличение мощности и скорости обработки данных, платформенные решения работы с данными смогут обеспечить обработку данных, их очистку, сопоставление, качество, полноту, публикацию, распространение практически в режиме реального времени, что позволит не только работать с ними онлайн, но и формировать отчеты и корректировать их с применением технологий искусственного интеллекта фактически на постоянной основе в режиме реального времени, формируя всевозможные срезы по данным по мере их появления и актуализации. В подобном режиме существенно повышается качество анализируемой информации и принятие управленческих решений, базирующихся на использовании невероятных массивов качественных данных, поражающих воображение их объемом. Соответственно, поскольку объемы данных являются основой и ценнейшим ресурсом использования самых передовых решений в области аналитики, то сама область применения и результаты исследований претерпит существенные изменения. Работа с данными на лету приведет к эволюции индустрии работы с данными, обеспечив единые правила работы как с основными данными, так и с транзакционными данными, оставив между ними весьма условную границу, проходящую исключительно через методологию работы с ними.

Как известно, инновационный процесс состоит из нескольких этапов: научные открытия, за которыми обычно стоят академические университеты и лаборатории; прикладные исследования, проходящие чаще всего в стенах корпораций и использующие в качестве основы академические открытия; этап изготовления изделий/продуктов; создание новых рынков сбыта. Это традиционный путь инноваций от их зарождения до возможности использования в повседневной жизни, и в случае аппаратного обеспечения эти этапы используют практически все инновации, попадающие на рынок. Любой существенный скачок развития аппаратного обеспечения, приводящий к кардинальному изменению производительности, обеспечивает мультипликативный эффект развития решений работы с данными.

В ближайшем будущем решения в области управления данными станут привычной областью во всех сферах деятельности, и одним из критериев оценки профессиональных качеств сотрудников станут компетенции, образование и опыт работы с информацией. Уже сейчас многие профессии хотя бы косвенно имеют дело с данными, формируя справки, отчеты, заполнения различных форм, и выполняя большинство повседневных операций за компьютером.

В будущем с учетом как существенного увеличения уже накопленных данных, так и роста поступающих в постоянном режиме данных от всевозрастающего количества систем и устройств, формирующих огромные массивы информации, можно ожидать отдельного направления экономики – организации, занимающиеся исключительно работой с данными с использованием различных платформенных решений для этого. Организации, основной деятельностью которых будет работа с данными, станут крайне востребованными в различных сегментах, где необходимо обеспечивать государственные и коммерческие организации инструментарием и услугами по формированию массивов данных, их сегментированию, обеспечению качества, полноты, непротиворечивости, интероперабельности, приведению к единому формату и многими другими услугами.

Более того, этот сегмент очень скоро будет самым востребованным на рынке в первую очередь за счет взрывного спроса со стороны многочисленных заказчиков, многие из которых не готовы заниматься этим направлением самостоятельно в силу дороговизны специалистов, инструментария, отсутствия компетенций и экономической нецелесообразности. По мере развития технологий в области искусственного интеллекта и их использования в платформах управления данными участие человека в

подавляющем большинстве процессов работы с данными будет снижаться, а эффективность решений, наоборот, увеличится.

Как мы неоднократно отмечали, мы живем в эпоху данных, которые нас окружают, даже если мы этого не замечаем, а разработки решений искусственного интеллекта базируются на четырех принципах:

- огромные объемы данных;
- вычислительные мощности;
- введение инструментария работы с данными;
- привлечение талантливых специалистов в области разработки алгоритмов работы с данными системы искусственного интеллекта.

Но основную роль играют именно данные, их объем и качество. Развитие синергии технологий работы с данными и систем искусственного интеллекта приведет к расширению функциональности платформ управления данными, в которых на единых принципах хранения данных будет реализовано подавляющее большинство известных сегодня отдельных решений, таких как MDM, DQ, DG, DATAMART, ETL, BI и других в единой платформе, позволяющей работать с основными и транзакционными данными, объемы которых будут исчисляться сотнями миллиардов записей. На рисунке 17.1 представлена продуктовая линейка одного из вендоров в области управления данными, учитывающая все вышесказанное.

Решения искусственного интеллекта, являющегося частью подобных платформенных решений, позволяют существенно улучшить качество данных, обеспечить управление данными, их мониторинг и операций с ними в зависимости от реализуемых сценариев, сегментируют информацию в соответствии с задачами, обеспечивают интеллектуальную настройку всевозможных правил и процедур в зависимости от многочисленных параметров данных, их структуры, внешних систем (источников и получателей данных); ранее используемых операций и все больше, все эти системы искусственного интеллекта могут принимать самостоятельные решения, формируя всевозможные подсказки в различных форматах.

На определенном этапе платформы управления данными, используя искусственный интеллект, обеспеченный большим объемом высококачественных данных, позволят принимать критически важные решения во всех сферах человеческой деятельности, окружая нас незаметной заботой и помощью – от управления всеми видами транспорта и проведения сложнейших хирургических операций до автономной работы большинства приложений на смартфонах и домашних устройствах.

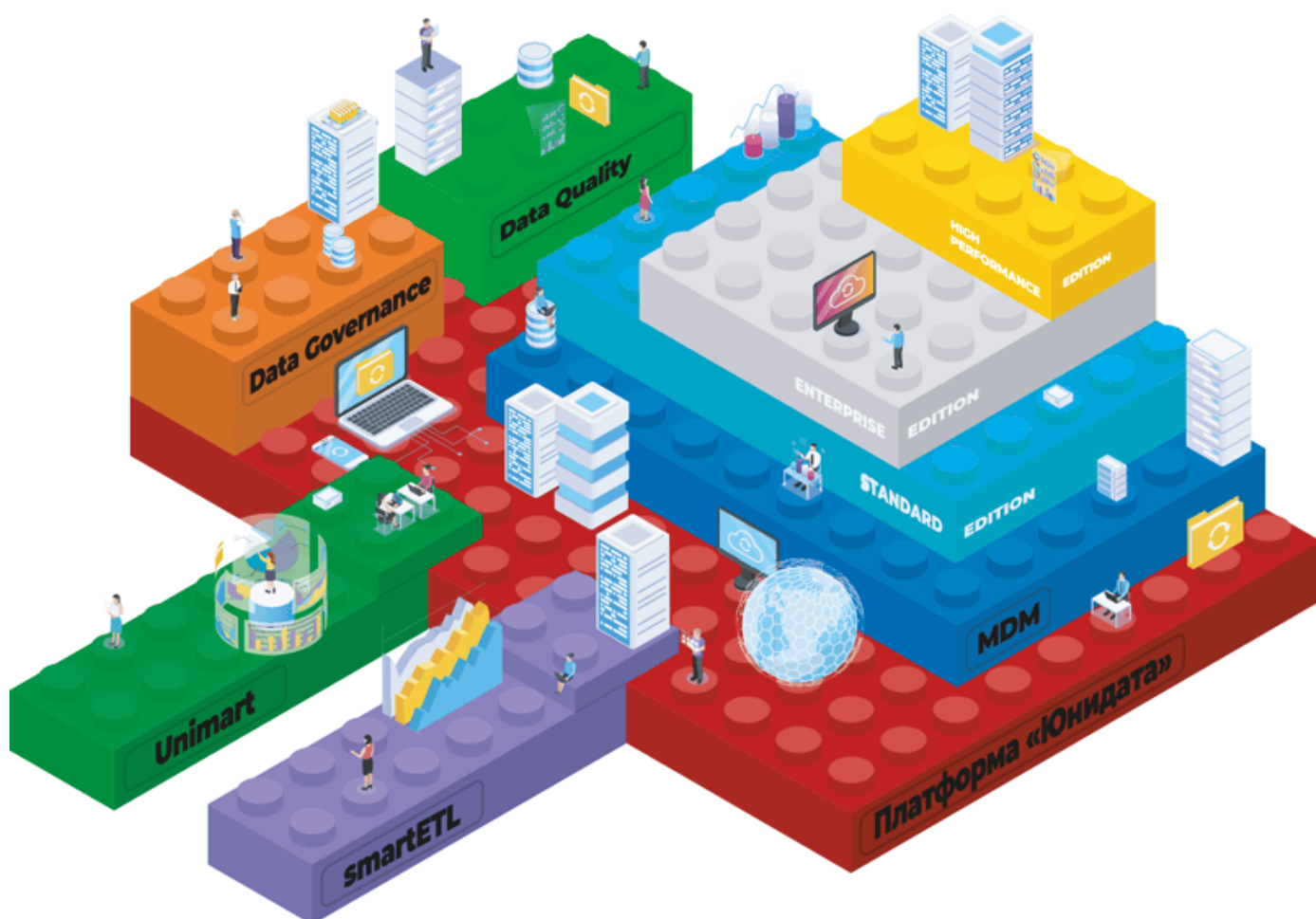


Рис. 17.1.

Продуктовая линейка одной из компаний, специализирующихся в области управления данными



ПРАКТИЧЕСКИЙ ПРИМЕР

Постепенное создание практики управления данными в «Телеком Дубль» дало компании возможность не только экономить на постоянном обновлении ИТ-ландшафта, но и получать от своих информационных ресурсов серьезную прибыль. Совмещение в одном месте данных по всем информационным доменам дает возможность получать аналитику по рынку телекоммуникационных услуг и с ее помощью развивать бизнес. Это развитие идет в том числе путем внедрения в собственную телекоммуникационную платформу новых видов продуктов – от курсов онлайн-обучения до медицинских консультаций.

Взглянув на бизнес глазами директора по данным, компания, в частности, увидела один из своих продуктов не просто как программу по организации видеоконференций, а как международную коммуникационную платформу с возможностью развития торговых площадок и привлечения новых бизнес-партнеров. И все это с использованием понятных и предсказуемых процессов.

В целом же можно отметить целый ряд полученных компанией очевидных выгод, например:

- серьезное увеличение показателей ARPU (average revenue per user – средняя выручка на одного пользователя) и LTV (customer lifetime value – общая прибыль, которую получают от

клиента в период сотрудничества);

- возможность за счет интеграции данных о клиенте в единый профиль и надстраивания над ним аналитики и системы принятия решений формировать оптимальное действие для клиента (next best action);

- возможность формировать максимально приближенные к реальности KPI для каждого сотрудника и т. д.

Желание «Телеком Дубль» постоянно развиваться, быть в тренде и задавать собственные правила игры позволило создать полноценную практику управления данными с достаточно высоким уровнем зрелости. Благодаря повышению экспертизы в области управления информационными активами из рядовой телекоммуникационной компании «Телеком Дубль» стала превращаться в одного из лидеров рынка, повышать капитализацию и играть все большую роль в отрасли.

Литература к главе 17

- Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2018.
- Anderson C. Creating a Data-Driven Organization. O'Reilly Media, 2015. (Русский перевод: Андерсон К. Аналитическая культура. От сбора данных до бизнес-результатов. – М.: Манн, Иванов и Фербер, 2017.)
- Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.
- Loshin D. Business Intelligence: The Savvy Manager's Guide: 2nd Edition. Morgan Kaufmann, 2012.
- Loshin D. Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph: 1st Edition. Morgan Kaufmann, 2013.
- Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need. Entity Group Limited, 2016.

Глава 18. Основные тренды будущего

В последней главе мы остановимся на основных трендах будущего в области управления данными. Важно отметить, что это не визионерский взгляд, а квинтэссенция того, о чем мы говорили на страницах нашей книги. Но перечисления последних трендов мало, мы пошли чуть дальше – рассказав об основных тенденциях (наука о данных, искусственный интеллект, машинное обучение и интернет вещей), мы покажем на ряде примеров, как управление данными влияет на вполне конкретные области человеческой жизнедеятельности – медицину, логистику, транспорт, сельское хозяйство, природопользование, промышленность, жилищно-коммунальное хозяйство. Таким образом, из «космоса» (дальние перспективы) мы вернемся назад в «атмосферу» (что будет использоваться в самое ближайшее время).

18.1. Основные тренды на ближайшую перспективу

Сейчас много внимания уделяется искусственному интеллекту (ИИ), причем в различных его проявлениях – от роботизации до машинного зрения и автоматизированного принятия решений. Ученые много лет пытаются научить машины «думать, как люди», в результате чего возникло машинное обучение, основанное на огромных массивах данных, которые нужно «должным образом подать и приготовить» для машины/алгоритма, чтобы она обучалась, а потом «подавать» еще, чтобы скорректировать ее работу и т. д. Это отдельная быстроразвивающаяся область, где сейчас особое внимание уделяется вопросам стабильности и адаптации алгоритмов машинного обучения к внезапным изменениям потоков данных, к возможным ошибкам данных или же нетипичным ситуациям, т. е. таким, к которым машину «не приучили». За примерами далеко ходить не надо – пандемия COVID-19 изменила многие устои, к которым мы привыкли (а значит, так и обучили машинные алгоритмы). Появилось большое количество отраслей, которые начали свою коренную перестройку. Об этих отраслях мы подробнее расскажем чуть дальше. И

здесь снова мы видим возрастающую роль данных – все уходит в онлайн, все больше сфер пытаются автоматизировать и роботизировать, чтобы избежать человеческого участия, все больше появляется массивов данных, доступных для различного анализа.

Интересный технологический тренд – распределенное хранение данных и параллельные вычисления. Все эти возрастающие массивы информации необходимо хранить, а вычислений становится больше. Понятно, что традиционные ИТ-гиганты выпускают больше дисков и систем хранения данных, способных вмещать все больше данных, а также еще более производительные процессоры, которыми наполняют мощные сервера. Но эта гонка не может продолжаться бесконечно. А как же все купленное три-пять лет назад? У многих из нас остались старые компьютеры и смартфоны. Именно здесь и сейчас наука серьезно занимается вопросами о том, как разместить данные децентрализованно и соединить их в нужный момент (здесь роль метаданных и управления ими особенно важна для распределенных хранилищ), как использовать простаивающие вычислительные мощности для необходимых вычислений и какие из них могут быть распараллелены с учетом нахождения данных для этих вычислений. Прослеживается четкая синергия с машинным обучением.

18.1.1. Дальнейшее развитие науки о данных

О науке о данных мы уже говорили. Это наука о методах анализа данных и извлечения из них ценной информации, знаний. Она тесно переплетена с такими областями как машинное обучение, наука о мышлении (cognitive science) и, конечно, с технологиями работы с большими данными. В свою очередь, большие данные – это огромные объемы неструктурированной информации: например, метеоданные за какой-то период, статистика запросов в поисковых системах, результаты спортивных состязаний, базы данных геномов микроорганизмов и многое другое. Для работы с такими данными используют математическую статистику и методы машинного обучения. Итог работы ученого по данным – прогнозная модель, некий программный алгоритм, который находит оптимальное решение поставленной задачи. Стоит отметить, что незаменимый помощник ученого по данным – специалист по машинному обучению, который выбирает архитектуру и обучающие алгоритмы.

В науке о данных слились воедино консервативные стратегические подходы, предметный анализ, приемы поиска шаблонов и аномалий данных, а также те алгоритмы поиска совпадений, нормализации, которые упоминались на страницах этой книги.

Наука о данных будет все больше использоваться в самых разных направлениях бизнеса. Приведем несколько примеров ее использования на практике. Для задач маркетинга эти технологии могут прогнозировать спрос на услуги компании. Для транспортных компаний – использоваться для выстраивания оптимального маршрута. Для производства – прогнозировать сроки выхода из строя оборудования и обнаружение дефектов продукции. В ИТ-сфере специалист в области науки о данных может разрабатывать ботов, поисковые алгоритмы, системы искусственного интеллекта. В сельском хозяйстве алгоритмы дают прогноз по урожаю, подбирают оптимальные системы землепользования. В страховом бизнесе – оценивают вероятность страхового случая. В медицине появляется все больше приборов, ставящих диагноз на основании данных, например, программа может указать поврежденные органы на рентгеновских снимках. Прогнозы погоды, которыми пользуемся, и те готовятся с помощью науки о данных.

Можно выделить три основных типа концепций, с которыми работает наука о данных⁵⁴⁸:

1. Представления о том, как наука о данных вписывается в организацию и конкурентную среду, включая способы привлечения, структурирования и развития команд, занимающихся наукой о данных; способы осмысления того, как наука о данных ведет к конкурентному преимуществу; и тактические концепции успешного выполнения проектов.

2. Общие способы анализа данных. Они помогают в определении соответствующих данных и рассматривают соответствующие методы. Эти концепции включают в себя процесс интеллектуального анализа данных, а также набор различных высокоуровневых задач интеллектуального анализа данных.

3. Общие представления о реальном извлечении знаний из данных, лежащие в основе огромного

⁵⁴⁸ Provost, Foster and Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, 2015.

массива задач науки о данных и их алгоритмов.

Фундаментальное понятие здесь – сходство двух объектов, описываемых данными. Способность определять сходство лежит в основе различных конкретных задач. Ее можно использовать, например, для поиска клиентов, похожих на данного клиента. Все это составляет основу нескольких алгоритмов прогнозирования, которые оценивают целевое значение, такое как ожидаемое использование ресурсов клиентом или вероятность того, что клиент ответит на предложение.

Сходство также составляет основу поиска информации, при котором извлекаются документы или веб-страницы, соответствующие поисковому запросу. Наконец, именно это понятие лежит в основе нескольких общих алгоритмов рекомендаций. Грубо говоря, когда мы посещаем интернет-магазин в поисках товаров и он рекомендует нам «посмотреть что-то еще», – это наука о данных в действии.



Рис. 18.1.

Наука о данных: основы

Здесь стоит привести очень интересный пример противостояния двух компаний-гигантов в области производства персональных компьютеров, Dell и Compaq, в 1990-е годы. К середине 1990-х Compaq была лидером американского рынка, но потеряла все, вовремя не отреагировав на угрозу со стороны Dell, которая использовала технологии науки о данных и получила важнейшие конкурентные преимущества. Речь идет о введении веб-сервиса по кастомизации компьютера, исходя из личных предпочтений заказчика. Compaq не смогла ответить еще и потому, что у нее была другая бизнес-стратегия продаж. В то время как Dell продавала PC напрямую через интернет, Compaq отгружала компьютеры через розничные магазины. И когда Compaq попыталась скопировать веб-стратегию Dell, она столкнулась с жесткой реакцией от своих розничных продавцов, потому что модель продаж не подходила для этой стратегии. Compaq вынесла из этой ситуации жесткий урок, который заключается в том, что на этапе понимания бизнеса нам необходимо тщательно подумать о том, как данные и наука о данных могут обеспечить ценность в контексте нашей бизнес-стратегии, а также будут ли они делать то же самое в контексте стратегий наших конкурентов. Через какое-то время Compaq стало лихорадить, некогда процветающий бренд купил другой конкурент, Hewlett-Packard, и вскоре после этого компания исчезла с рынка.

Есть и другой громкий пример – противостояние Amazon и Borders. Когда-то обе сети конкурировали в области книжных продаж, но введение Amazon персонализированных рекомендаций отправило Borders на покой. О компании сейчас никто даже не вспомнит, а детище Джеффа Безоса – одна из крупнейших ИТ-компаний в мире.

18.1.2. Огромный скачок в области искусственного интеллекта и машинного обучения

Технологии искусственного интеллекта (ИИ) – это важнейший инструмент работы с данными. Именно искусственный интеллект позволяет преобразовывать огромный объем данных в различные сервисы и продукты, быстрота, удобство и качество которых в первую очередь зависят от объема и качества используемых данных и применяемых технологий искусственного интеллекта.

Если обратиться к истокам ИИ, то первые упоминания о нем относятся к 1950-м годам, когда ученые продемонстрировали начальные версии работы искусственных нейронных сетей. К сожалению, после первых успехов исследования в этой области довольно быстро остановились, в первую очередь из-за недостаточности данных и вычислительных ограничений аппаратного обеспечения того времени.

В дальнейшем исследование ИИ шло эволюционным путем. Сначала речь шла о создании решений на основе правил, затем – на принципах нейронных сетей. В основе нейронных сетей также лежали простейшие правила, которые задают направление движения «мысли системы». В отличие от экспертных систем, нейронные сети обучаются не на правилах, а на текстовых входных и выходных данных, имитирующих правила обучения человеческого мозга. Решения на основе правил базировались на опросе экспертов, а их ответы составляли суть решений, называемых также экспертными, логику которых можно описать правилом «если-то». У экспертных систем существует несколько слабых мест, главное из которых – невозможность эффективного использования этого метода при существенном расширении задач и предметных областей.

Когда исследуемые системы стали сложнее, правила стали складываться в конечные автоматы, т. е. последовательности применения правил в зависимости от исхода предыдущего шага. Именно тогда появились нейронные сети. Исследователи-сторонники нейронных сетей не задают правил, основанных на ответах экспертов. Нейронные сети – это шаг в сторону имитации работы человеческого мозга, т. е. это метод использования огромного массива образцов картинок, звуков, математических решений, видеофрагментов для самостоятельного определения закономерностей загруженных данных, используя правила обучения человеческого мозга. В качестве примера можно привести обучение ребенка, когда он видит впервые в жизни какой-то предмет или животное и родители дают определение или другими словами отвечают на вопрос «Что это?». В дальнейшем ребенок, столкнувшись с этим предметом, уже знает из ответов родителей, что это такое, и в процессе взросления получает все больше описательной информации об этом предмете и самостоятельно оценивает увиденное/услышанное, делает определенные выводы и принимает самостоятельные решения.



Рис. 18.2.

Искусственный интеллект, машинное обучение и их «соседи»

По мере усложнения задач развитие технологий нейронных сетей требует не только существенного развития вычислительных мощностей, но и огромного объема данных, а также алгоритмов обучения все увеличивающихся слоев искусственных нейронов. Во многом благодаря исследованиям Джеффри Хинтона⁵⁴⁹ в 2000-х годах появился термин «глубокое обучение», основанный на открытом им способе эффективного обучения добавленных слоев.

Перспективы технологии нейронных сетей привели к тому, что решения ИИ, основывающиеся ранее на экспертных знаниях, теперь целиком и полностью базируются на данных. По мнению одного из известнейших мировых экспертов в области искусственного интеллекта Кай-Фу Ли, «для создания эффективных алгоритмов искусственного интеллекта нужны три составляющие: большие объемы данных, вычислительные мощности и труд способных – но не обязательно выдающихся – разработчиков алгоритмов искусственного интеллекта»⁵⁵⁰.

При этом основу развития технологий искусственного интеллекта сейчас составляют данные, их объемы и качество. Чем выше значения этих параметров, тем более эффективные алгоритмы они позволяют создать, независимо от направлений деятельности этих решений. Преимущество в технологиях ИИ получают те страны и компании, у которых формируется наибольший объем данных, которые из количества переходят в качество. ИТ-гиганты многие годы накапливают данные пользователей и их активности – как в онлайн-платформах (различные транзакции, приобретаемые услуги и товары, их характеристики, история поисков, загруженные и просмотренные фотографии, видео, комментарии и лайки

⁵⁴⁹ Hinton G., LeCun Y. Deep Learning, Nature Publishing Group, 2022.

⁵⁵⁰ Ли Кай-Фу. Сверхдержавы искусственного интеллекта. Китай, Кремниевая долина и новый мировой порядок / Кай-Фу Ли; пер. с англ. Нины Константиновой. – М.: Манн, Иванов и Фербер, 2018.

и прочие активности), так и данные из реальной жизни (банковские транзакции, использование различных финансовых и страховых сервисов, посещение ресторанов, фитнес-клубов, и данные о разнообразных пристрастиях, данные о путешествиях и многих других данных, собираемых бизнесом не только о пользователях, но и о членах их семей, друзьях и коллегах). Все это позволяет собрать полноценную картину повседневной жизни граждан и бизнеса.

По мнению Кай-Фу Ли, «полный переход к повсеместному использованию искусственного интеллекта займет некоторое время и будет состоять из четырех волн – т.е. этапов внедрения искусственного интеллекта. Первым мир покорит ИИ интернета, затем ИИ для бизнеса, потом наступит черед ИИ восприятия и автономного ИИ. На каждом из этих этапов ИИ будет захватывать новые области нашей повседневной жизни»⁵⁵¹.

В настоящее время *ИИ интернета* и для бизнеса уже ощущается во многих сферах нашей жизни. Мы получаем различные советы от поисковых систем, видим на страницах браузеров рекламу и описание интересующих нас товаров и услуг, зачастую расположенных рядом с нашим домом и офисом, читаем новости, подобранные с учетом наших интересов. Все это настолько органично вошло в нашу жизнь, что без этого сложно представить использование интернета современным человеком.

В свою очередь *ИИ для бизнеса* уже стал неотъемлемой частью многих бизнес-процессов, заменяя собой многие алгоритмы работы организации, автоматически формируя и проверяя документы, оценивая заемщиков, планируя маркетинговые акции и эффекты от их проведения, формируя тарифные планы и экономические результаты их возможного внедрения, диагностируя болезни и рекомендации по лечению, оценивая результаты различных фармакологических исследований и возможностей распространения тех или иных болезней, и многое другое.

ИИ восприятия фактически переводит в цифровой вид окружающий нас мир; все, что человек слышит, читает, воспринимает. Все это уже в большинстве своем оцифровано, и благодаря дальнейшему развитию технологий этот объем цифровой информации будет кратно возрастать с каждым годом, учитывая широкое распространение датчиков, камер, и других устройств. Все данные, поступающие с огромного количества телефонов, телевизоров, автомобилей, камер наблюдения, различных приложений, постоянно анализируются и формируются в кластеры, позволяющие строить цифровую среду с небывалой скоростью и качеством, неотличимым, а где-то превосходящим наш физический мир, делая цифровой мир комфортнее и обогащая его все новыми возможностями.

Автономный ИИ – венец развития цифрового мира. Он обеспечит интеграцию достижений предыдущих трех волн, на основе огромного массива собранных данных и мгновенной обработки текущей информации со всевозможных устройств. Автономный ИИ кардинально изменит мир, к которому мы привыкли. Отдельные решения уже становятся частью нашей жизни, это касается прежде всего транспортной среды (беспилотные автомобили), различные технологии розничной торговли (беспилотные кассы и вендинговые аппараты), логистики и складского обслуживания (беспилотные погрузочные и сортировочные машины), сельскохозяйственная техника, грузовые автоперевозки и многие другие решения.

В настоящий момент на рынке решений ИИ доминируют Google, Amazon, Microsoft, Baidu, Alibaba и Tencent. Все они, кроме Baidu, входят в десятку крупнейших компаний мира по рыночной капитализации, что указывает на беспрецедентный рост капитализаций компаний, работающих в области ИИ, отношение к единому направлению со стороны инвесторов и перспективность дальнейшего развития и монетизации технологий. Эти компании аккумулируют гигантские суммы в исследования ИИ. А если учитывать накопленные ими гигантские объемы данных в самых разнообразных сферах и, как следствие, привлечение лучших специалистов, то станет понятно, что эти компании фактически недостижимы для других игроков рынка. Кроме того, с каждым днем внедряются все новые технологии, количество и качество данных растут, что создает еще больший задел. Как мы говорили ранее, для дальнейшего развития технологий ИИ самое важное – это объемы и качество данных, что позволяет на их основе проводить новые исследования и создавать инновационные продукты.

Основными направлениями применения ИИ становятся различные отрасли бизнеса, которые накладывают свои требования к применяемым технологиям. Например, машинное обучение, о котором

551 Там же.

пойдет речь далее, развивается в машинное зрение, которое широко используется в телемедицине и автостроении. Другое применение находят ИИ технологии в предиктивной аналитике, где ранее были исключительно статистические методы. Технологии интернета вещей, о которых пойдет речь далее, также базируются на ИИ, особенно в т. н. индустриальном интернете вещей (IIoT).

Существуют три классических модели внедрения технологий ИИ для компаний.

1. Новые форматы сбора и информации о клиентах/продуктах/сервисах и возможностях взаимодействия с ними.
2. Возможность предложения новых продуктов/сервисах.
3. Совершенствование бизнес-процессов и автоматизация производства.

Если говорить об управлении данными, то важно отметить, что информация из одного источника должна обогащать данные, полученные из другого. Это, в свою очередь, ведет к синергии, но также может приводить и к конфликтам данных и созданию ошибок. Например, при объединении данных огромной размерности могут наступить фатальные последствия из-за невозможности внесения исправлений в больших масштабах.

Машинное обучение – одна из областей ИИ, которая состоит из методов, позволяющих компьютерам принимать решения на основе данных и реализовывать приложения ИИ. Фактически, машинное обучение – класс методов ИИ, который ориентирован на создание систем, которые обучаются (т. е. совершенствуют свою производительность) посредством анализа данных. Понятия «машинное обучение» и «ИИ» часто употребляют в одном контексте порой как взаимозаменяемые, однако они имеют разное значение. Разница состоит в том, что машинное обучение всегда подразумевает использование ИИ, однако ИИ не всегда подразумевает машинное обучение.

Машинное обучение присутствует практически во всех сферах современной жизни. Каждый раз, когда мы пользуемся банковскими услугами, делаем покупки в интернете или общаемся в мессенджерах, алгоритмы машинного обучения помогают сделать это взаимодействие удобнее, эффективнее и безопаснее.

В основе машинного обучения лежат алгоритмы, которые разделяются на два основных типа: обучение «с учителем» и без него. Разница заключается в способе изучения данных для последующего прогнозирования.

Практически все мировые технологические лидеры уже приняли стратегию разработки своих продуктов и сервисов по так называемой дата-центричной модели. Она подразумевает построение новых продуктов и сервисов, исходя из имеющихся данных, которыми эти решения будут пользоваться, и данных, которые эти решения будут формировать, обогащая ими текущие информационные массивы и «цифровые профили».

Такие компании, как Alphabet и Google, считают, что ИИ – основа для революции компьютерных технологий, и его влияние на общество будет даже более значительным, чем появление интернета. Крупнейший производитель напитков Coca-Cola в день продает более 1,9 млрд товарных единиц. У компании более 500 брендов. Приведем цитату руководителя производства Coca-Cola Грега Чемберса: «Искусственный интеллект – фундамент всей нашей деятельности. Мы создаем интеллектуальный опыт, и его ядро – это искусственный интеллект»⁵⁵².

В качестве еще одного интересного примера можно привести платформу SmartThings компании Samsung, которая может стать первой компанией в мире, реализовавшей концепцию «сетевого (цифрового) дома», в рамках которой все устройства Samsung смогут обмениваться данными в единой цифровой экосистеме. Все это открывает поистине безграничные возможности для различных новых сервисов и формата взаимодействия с потребителями на уровне проектов и сервисов. В этой связи можно констатировать что любой разработчик цифровой экосистемы, получившей широкой распространение и ставшей неким стандартом объединения и работы с данными/продуктами/сервисами, получит огромное преимущество в будущем.

18.1.3. Развитие интернета вещей

⁵⁵² How Coca-Cola and PepsiCo Use AI To Bubble Up Innovation. – URL: <https://analyticsindiamag.com/how-coca-cola-and-pepsico-use-ai-to-bubble-up-innovation/3>

Интернет плотно вошел и в нашу жизнь, и в окружающую нас действительность. Это не только компьютеры, смартфоны и планшеты, с помощью которых мы «выходим в интернет». Речь идет практически обо всей бытовой технике. Холодильники считывают сроки годности и составы продуктов, пылесосы и метеодатчики работают в доме в наше отсутствие, современные автомобили считывают всевозможные показатели, от трафика до метеоусловий, а «проникновение» интернета в медицинские приборы вызывает настоящее восхищение. Все эти предметы генерируют и собирают данные, причем объем этих данных поистине огромный. Огромное количество приложений обрабатывают эти данные и генерируют новые: холодильники подсказывают, когда у каких продуктов истечет срок годности, состав всего загруженного в них, часы обладают множеством функций, подсказывающих сколько шагов, калорий уже потрачено, когда необходимо пить воду. Автомобили пользуются машинным зрением – сами паркуются, поворачивают и тормозят. Все эти функции были бы невозможны без надлежащей обработки информации, трактовки исторических данных и соответствующей синхронизации и обмена информацией по каналам всемирной сети интернет.

Во всех этих примерах важно отметить различные аспекты работы или, здесь уже можно сказать, управления данными, которые мы уже успели рассмотреть в этой книге. Действительно, все начинается с «понимания» (семантики) данных – холодильник должен понять «что» в него ставят, при этом речь не идет о стандартных бирках – могут быть разные магазины, продуктовые рынки и даже ручная корректировка вводимых данных. Вопрос качества данных безусловно очень важен, ведь если умные часы собирают геоданные с ошибкой, то все приложения, использующие эту информацию, будут давать неверные результаты скорости движения, геопозиционирования и пр. Ну и конечно, функции мониторинга потоков данных со всевозможных датчиков, «раскладки этих данных» согласно семантике и метаданных являются основополагающими для всех этих приложений трекеров, смартфонов, телемедицины, автомобильной, спортивной и прочих отраслей.

Все это были примеры из нашей бытовой жизни (Commodity Internet of Things). А что происходит в производстве, промышленности? Конечно, там все началось гораздо раньше с автоматизации производственных процессов, создания специализированных программ проектирования, управления станками и практически всем оборудованием. Сюда также пришел интернет вещей, его называли промышленным интернетом (IIoT, Industrial Internet of Things). Попробуем описать его специфику на двух примерах. Первый приведет нас к образу «Цифрового двойника» из фантастических романов начала XX века, а второй – ситуация, которая полностью поменяла бизнес-модель целой компании.

Первый пример – крупная международная компания, которая занимала (и продолжает занимать) одно из лидирующих мест по производству турбин и промышленных буровых установок (здесь и далее мы чуть упрощаем терминологию для широкой аудитории наших читателей). Конечно же, производство модернизируется, а турбины оснащаются все большим количеством датчиков (или сенсоров), снимающих такие показатели, как отклонение от вертикальной оси, западение скорости вращения и др. В определенный момент в компании осознали, что объем датчиков покрывает полностью все возможные состояния оборудования как в статичном положении, так и в динамике. Полученный образ можно назвать полным «цифровым двойником» и отслеживать любые его состояния удаленно. Кроме того, все оборудование, которое взаимодействовало друг с другом, оказалось оснащено аналогичными и подобными датчиками. Благодаря этому было решено сразу две задачи: во-первых, все турбины, управляющие установки и прочее оборудование оказались полностью оцифрованы, имели своих «цифровых двойников», которые получали информацию от соответствующих датчиков, а во-вторых, получилась единая и полностью интегрированная экосистема цифровых двойников» за счет единообразия протоколов обмена информации между датчиками. Следующим шагом компания стандартизировала интерфейсы обмена информации для всех датчиков всех узлов своего оборудования, поскольку большую часть узлов они закупали у партнеров, а сами занимались сборкой.

К чему это привело? Цифровые двойники» сильно упростили сборочный процесс, отбраковку узлов. У компании появились дополнительные услуги (ИТ-сервисы), например:

- постоянный мониторинг оборудования за счет цифровых двойников, что сокращает затраты на персонал, ответственный за это оборудование;
- предупреждение аварийных ситуаций, улучшение техники безопасности за счет постоянного считывания информации с датчиков и возможность моделирования всевозможных ситуаций на цифровых

двойниках;

- услуги по предикативной/предсказательной аналитике, что важно не только для предупреждения аварий, но также позволяет своевременно заказать запасные части или сервисное обслуживание данного оборудования.

Список дополнительных сервисов далеко не полный, и мы уверены, эта компания продолжает их добавлять, оставаясь лидером отрасли. Именно этот шаг больше закрепил лидерские позиции, поскольку одним из основных достижений введения единой платформы IIoT и экосистемы стало привлечение еще большего числа партнеров-производителей более мелких узлов оборудования, т. к. им стало проще работать на уровне цифровых двойников и взаимодействовать с реальными конструкторскими образцами уже на последних этапах.

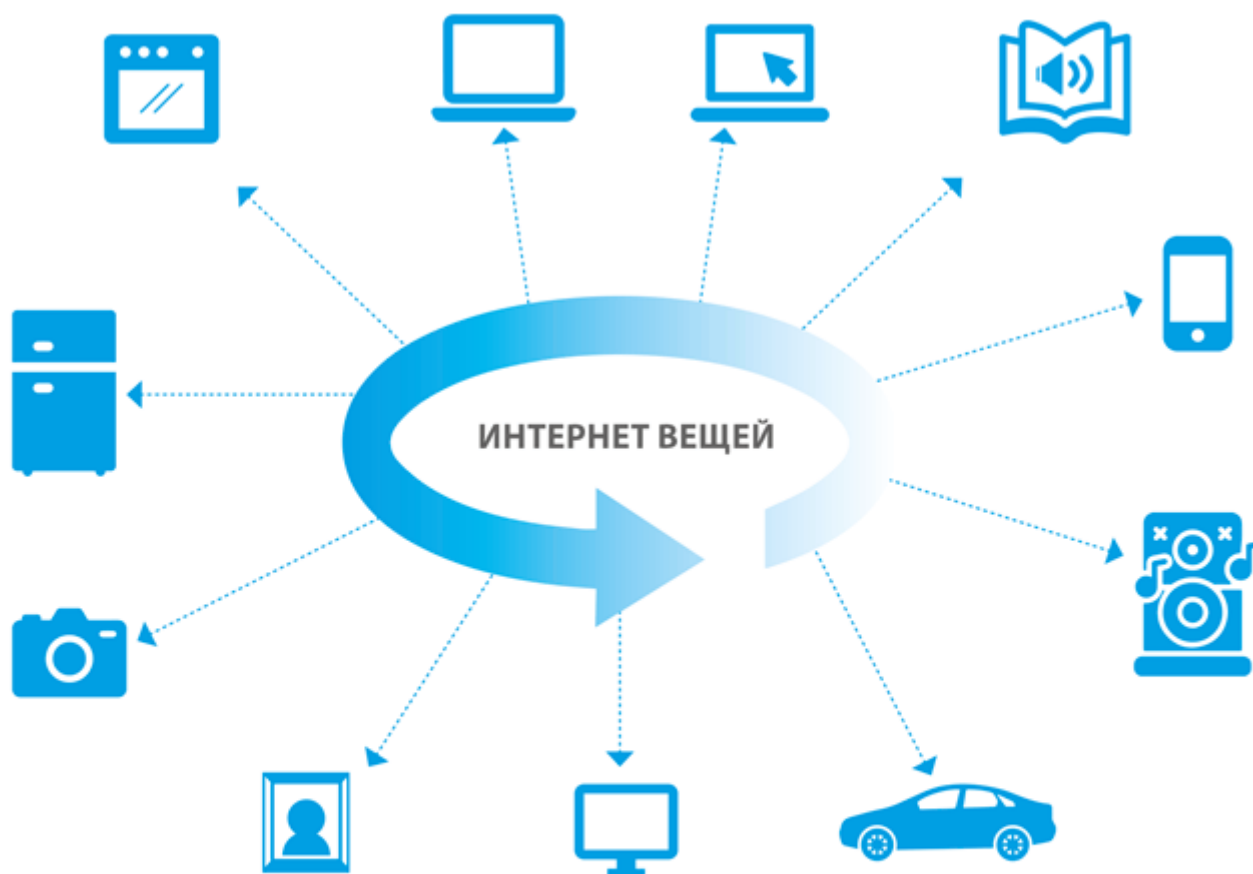


Рис. 18.3.

Бытовые приборы, объединенные в единый интернет вещей

Вторым ярким примером использования концепции IIoT, а в этом случае правильной обработки данных от всевозможных датчиков своих изделий, стала компания GE Aviation. Это один из лидеров производства авиационных двигателей, занимающий одну из первых строчек наряду с CFM, Pratt&Whitney и Rolls Royce. Авторы не были вовлечены во весь цикл цифровизации авиационных двигателей в отличие от первого примера. Однако к моменту знакомства с проблематикой управления данными было уже очевидно, что «цифровые двойники» созданы и все перечисленные сервисы мониторинга, предиктивной аналитики и пр. существуют.

В данном примере компания GE Aviation пошла дальше и, благодаря наличию сервисов, смогла предложить и в большей части перейти к сервисно-ориентированной экономике. Что стоит за этими словами? Все просто: теперь можно прийти в компанию и купить не авиационный двигатель, а часы налета этого двигателя. Двигатель привезут, произойдет монтаж. А далее в специальном авиационном центре GE

Aviation будут отслеживать его работу, предупреждать внештатные ситуации. Сами своевременно закажут необходимые запчасти, проведут сервисные работы, что не приведет к тому, что двигатель будет простаивать, ожидая ремонт, а заказчик и GE Aviation будут недополучать работу двигателя и деньги за обслуживание. Эта концепция оказалась взаимовыгодной и востребованной, и при этом была бы невозможна без экосистемы датчиков, объединенных единым интерфейсом обмена данными. Это и есть яркий пример, когда современный IoT превращает консервативные и устоявшиеся компании в современные конгломераты с сервисно-ориентированной экономикой.

18.1.4. Формирование цифровой эры постиндустриальной экономики

Следующий тренд – формирование цифровой эры постиндустриальной экономики. И что самое любопытное, существенные изменения в жизнях миллиардов людей в связи с эпидемией COVID-19 лишь ускорило ее формирование. Одним из результатов COVID стала огромная государственная поддержка таких перспективных направлений деятельности человека, как фармацевтика и биотехнологии. Глобальное развитие этих отраслей традиционно передвигали «на будущее». И коронавирус фактически и создал это «будущее».

Приблизил эпидемия и технические возможности по организации удаленного труда. Мы имеем в виду информационные технологии и различные сервисы, являющиеся частью ИТ, которые позволяют дистанционно работать, заказывать товары и услуги, удаленно использовать вычислительные мощности, находящиеся за тысячи километров, а также возможности по использованию всевозможных образовательных, медицинских сервисов, возможности приобретения различных товаров не выходя из дома.

К людям пришло осознание, что ускоренное развитие технологий возможно исключительно на базе уже имеющихся решений, которые необходимо переводить в цифровой формат. Без современной технологической основы в различных областях (машиностроение, автомобилестроение, энергетика, медицина, телекоммуникации и многие другие) создать цифровые технологии невозможно, поскольку отсутствует технологическая основа для цифровизации. Увы, при всех достижениях современных технологий область работы с данными остается крайне недооцененной, поскольку все современные технологии эпохи post-COVID опираются на огромные массивы данных, и роль данных будет только возрастать.

Выдающийся немецкий физик, нобелевский лауреат, основоположник квантовой теории Макс Планк говорил: «Не следует думать, что новые идеи побеждают путем острых дискуссий, в которых создатели нового переубеждают своих оппонентов. Старые идеи уступают новым таким образом, что носители старого умирают, а новое поколение воспитывается в новых идеях, воспринимая их как нечто само собой разумеющееся»⁵⁵³.

Слова Планка актуальны как никогда: сегодня активно формируется новая технологическая эра, характеризующаяся стремительным развитием управления данными. Это уже не просто ИТ, а новая, следующая ступень технологического развития. Не просто информационные технологии, а технологии управления данными, цифровая эра постиндустриальной экономики.

Стремительное развитие технологий имеет серьезные последствия для персонала компаний, стремящихся стать технологическими лидерами: внедрение новейших технологий работы с данными, включая искусственный интеллект и машинное обучение, по оценкам некоторых специалистов приведут к существенному сокращению персонала (до 80 % в последующие десять лет). Все это потребует переобучения практически всех специалистов, поскольку знание технологий работы с данными станет такой же базовой необходимостью, как сейчас – знание офисных программных продуктов. В то же время для привлечения ценных специалистов компании должны заслужить репутацию технологического лидера, что потребует существенных инвестиций в развитие технологий и кардинальное изменение модели привлечения специалистов.

Технологические лидеры доковидной эпохи, получившие технологические преимущества в постиндустриальную эпоху, сейчас становятся с технологической точки зрения недостижимыми и при переходе в цифровую эру.

553 M. Planck. Treatise on Thermodynamics 3rd English ed., A. Ogg, trans., Dover, New York, 1944.

В качестве примера можно рассмотреть телекоммуникационную отрасль (вспомним нашу абстрактную компанию «Телеком Дубль» из предыдущих глав), которая еще 20–30 лет назад развивалась семимильными шагами: велись огромные проекты по построению волоконно-оптических линий связи, модернизировалось оборудование коммутации (для перехода к пакетной коммутации), создавались огромные центры обработки данных для хранения различной информации, развертывались сети мобильной связи целых регионов и стран. Все это позволило довести показатель проникновения сетей связи до абсолютных значений, но основной задачей того периода являлось предоставление качественной голосовой связи независимо от географии абонента.

В начале 2000-х годов сначала руководители ИТ-корпораций, а потом и топ-менеджеры компаний в области телекома стали осторожно говорить, что построенная инфраструктура гораздо «больше» чем то, что нужно просто для предоставления голосовых услуг. Появились идеи на базе уже построенных сетей предоставлять принципиально новые сервисы (тогда это называлось «услуги с высокой добавленной стоимостью»). Именно в этот период стали появляться первые смартфоны, позволяющие использовать ИТ-сервисы. Отметим, что хотя те смартфоны сильно отличались от того, к чему мы привыкли сегодня, они позволяли обмениваться сообщениями, фотографиями, просматривать сайты, заказывать товары и услуги, а также, пускай и с определенными ограничениями, работать с почтой.



Рис. 18.4.

Цифровая эра постиндустриальной экономики

Именно в тот период нынешние технологические лидеры сформулировали свое видение телекоммуникационного будущего: какое решение ближе всего к абоненту, то решение и будет иметь преимущество, сети связи станут просто трубой для доставки данных. Это и произошло в последующие 10–15 лет. Компании, первыми это осознавшие и инвестировавшие в решения для абонентов или даже в «новую реальность или экосистему» для абонентов, такие как Apple, Google, Microsoft, Samsung, Netflix, Zoom и многие другие, стали мировыми технологическими лидерами, капитализация которых на порядки

превышает капитализацию телекоммуникационных гигантов начала 2000-х годов.

Все это стало возможно благодаря синергии, позволяющей смело смотреть в будущее развитие технологий, а также из-за изменения поведенческих моделей современного общества, возможности привлечения существенных инвестиций и предпринимательской смелости руководства компаний – технологических лидеров.

Стоит отдельно отметить, что почивать на лаврах свежесозданных технологий долго не получится – технологии сегодня сменяются в некоторых сферах каждые три-пять лет, и то и чаще. Как ни странно это может прозвучать, COVID открыл окно возможностей для огромного технологического рывка, позволяющего человечеству становиться еще более «цифровым», создавать новые платформы и сервисы, переходить на новые технологии практически во всех направлениях. Физическое местонахождение людей становится все менее важным, что неминуемо приведет к развитию функционала так называемых цифровых двойников. И если в нашем примере выше об интернете вещей речь шла о цифровых двойниках деталей и агрегатов, то сейчас можно говорить уже о двойниках людей.

Все больше игроков рынка в разных областях трансформируются в технологические компании. Более того, большинство новейших технологий, необходимых для внедрения новых продуктов и сервисов, а также для ускорения их успешного вывода на рынок, базируются на использовании данных как фундаментальной основе развития новых классов технологий. Именно поэтому качество данных становится основным приоритетом цифровой трансформации. На первый план выходят такие показатели, как правильность, полнота, непротиворечивость данных, возможность использования единых форматов данных, и интероперабельность, о чем мы подробно рассказывали в соответствующих главах книги. В свою очередь, формирование единого полного цифрового профиля человека, продукта, компании становится первоочередной и приоритетной целью компаний, планирующих свое технологическое развитие и стремящихся оставаться лидерами рынка.

Вместе с тем самые современные и востребованные технологии, такие как ИИ, машинное обучение, – это прежде всего технологии работы с данными. Их эффективность во многом зависит от качества и полноты используемых данных. Данные – это сырье для ИИ и машинного обучения. Практически все устройства, которыми мы пользуемся, собирают и передают данные, оставляя так называемый цифровой след.

Принципиально важное значение приобретают основные данные, которые корректируются и обновляются крайне редко – фамилия, имя, отчество, реквизиты документов, имущество, номер телефона и других устройств, реквизиты юридического лица, наименование и описание продуктов и сервисов и другие данные. Это так называемые центры притяжения/сбора транзакционных данных и основные параметры для их дальнейшего применения и аналитики.

18.2. Ценность данных в различных областях

18.2.1. Медицина и спорт

Современный мир стремительно меняется, и эти изменения в первую очередь связаны с развитием технологий во всех сферах жизни, прежде всего технологий управления данными. Сегодня для исследователей открыты возможности для сбора и анализа гигантских объемов данных, их систематизации, формирования статистических и аналитических отчетов, выявления различных тенденций, мониторинга изменений, создания прогнозов и проектов будущего развития как отдельных показателей, так и целых отраслей, что открывает невероятные возможности как для фундаментальной науки, так и для прикладных исследований. Отдельно хотелось бы отметить изменения, происходящие в одной из самых важных сфер жизнедеятельности человечества, а именно в области медицины.

Именно в медицине особую роль играет сбор достоверных данных из разных источников (будь то медицинское оборудование или услуги), отслеживание истории и версионности и получение в итоге полной и объективной картины пациента, основанной на актуальных и достоверных данных, которое позволит сделать правильные выводы и поставить диагноз. Именно здесь цена ошибки велика как никогда. И именно в этой области постоянно появляются новые инструменты работы с данными и новые научные направления.

В самое ближайшее время сформируется новая индустрия использования данных. Она вберет в себя

сразу несколько направлений: медицину, массовый спорт, фармацевтику и правильное питание. Данные станут объединяющим звеном всех этих направлений жизнедеятельности, а возможность сбора огромного массива информации позволит научному сообществу рассматривать все эти направления как взаимосвязанные. Эта сторона жизни человечества под влиянием технологий изменится самым кардинальным образом и преобразуется в отдельный сегмент жизни, который станет самым капиталоемким и инвестиционно-привлекательным на долгие годы.

Возвращаясь к технологиям, можно констатировать, что уже сейчас размываются границы между практической наукой, лечением, спортом, питанием и развлечениями. Новые технологии и объемы данных, позволяющие анализировать взаимосвязанные причины и следствия, уже сейчас формируют открытия, которые были ранее недоступны в силу невозможности настолько масштабных исследований. Мы стоим на рубеже научно-технологического взрыва в этой области. Отдельно отметим, что точность этих данных проверена до сотых долей значений, поскольку информация формируется датчиками.

В качестве подтверждения будущих открытий можно констатировать, что сейчас многие компании работают над созданием новых видов датчиков, позволяющих снимать различные показатели жизнедеятельности человеческого организма. Все большее внимание уделяется носимым устройствам, которые способны заряжаться от человека, т. е. трансформировать тепло человеческого тела в электроэнергию. Успех подобных технологий позволит уже в ближайшее время создать носимые датчики и гаджеты, которые никогда не придется заряжать. Все эти устройства будут круглосуточно сопровождать своего хозяина.

Подобные технологии также будут востребованы и в нейроимплантации. Вживленные в организм человека микрочипы и электронные устройства не придется постоянно подзаряжать или извлекать хирургическим путем для замены батареек. Эти устройства будут передавать необходимые показатели пациента и реагировать на внештатные значения, требующие медицинской помощи без вовлечения самого пациента, что крайне важно, когда речь идет о пожилых людях или об отсутствии рядом постоянного квалифицированного медицинского персонала.

Уже сейчас заметно, что предсказания ученых об изменениях в научном ландшафте приобрели гигантский характер: отдельные узкие специализации во многих направлениях исследований становятся частями сложных комплексных исследований, открытия которых зачастую появляются на стыке различных научных направлений: биологии, химии, физики, медицины, робототехники, различных ИТ-технологий, а в последнее время еще и различных технологий работы с данными. Уже сегодня международные и национальные организации в области здравоохранения вынуждены вести огромное количество реестров: международный и национальные классификаторы болезней, перечень лекарственных средств, перечень наименования лекарств, перечень активных веществ медицинских препаратов, перечень медицинского оборудования и их производителей, наличие многочисленных сертификатов, перечень медицинских работников и медицинских учреждений. А кроме того, требуется еще и целый пласт данных, формируемых на основе медицинских препаратов и оборудования, результатов медицинских исследований и анализов, и огромный пласт транзакционных данных от датчиков и изменений как периодических, так и в режиме реального времени.

Учитывая отсутствие технологических проблем в части сбора различной медицинской информации, ее систематизации и сопоставления, выявления тенденций и прогностических моделей, данные в этой области стали на порядки более точными. Все это позволяет не только проводить исследования на современном качественном уровне, но и существенно упростить и улучшить работу медицинских работников и учреждений, а также существенно повысить качество жизни многих пациентов, ставить более точные диагнозы, своевременно и точно обеспечивать и корректировать протоколы лечения, что, в свою очередь, спасает многие жизни. Технологии работы с данными также снимают географические границы для людей, позволяют им получить качественную медицинскую помощь в любой точке мира.

Наличие подобных технологий также открывает безграничные возможности для многих направлений бизнеса, так или иначе связанных с качеством жизни, таких как страховые компании, пенсионные фонды, банки, транспортные компании, аптечные сети и фармацевтические компании. Отдельно стоит отметить необходимость использования этих данных для различных государственных органов власти, министерств и ведомств, позволяющих формировать социальную политику, планы в области здравоохранения, различных видов страхования, пенсионного обеспечения, фармацевтики, обеспечения различных видов

государственной поддержки разных слоев населения, требования в области профессиональной деятельности, финансирования научных исследований и многих других направлений государственной политики. Вместе с тем использование технологий работы с данными существенно повысит требования к знаниям в этой области практически ко всем специалистам в сфере здравоохранения, а также к большинству сотрудников из смежных сфер профессиональной деятельности.

С другой стороны, в самом ближайшем будущем объем обрабатываемых данных позволит широко использовать технологии искусственного интеллекта, поскольку подобные объемы данных человек попросту не может объективно анализировать. Все это приведет к более точным и выверенным решениям и постепенно снизит роль человеческого фактора. Многие из того, что сейчас является прерогативой медицинских работников (постановка диагноза, назначение лечения, интерпретация результатов исследований и пр.), станут стандартными функциями медицинских комплексных систем, основанных на технологиях работы с данными. В дальнейшем подобные медицинские комплексы позволят не только заменить докторов в части назначения лечения и оценки результатов такого лечения, но и с развитием робототехники проводить сложнейшие хирургические операции. Все это изменит сферу здравоохранения до неузнаваемости, а медицинские работники вынуждены будут наряду с медицинскими специальностями осваивать новые для себя профессиональные навыки и специальности, связанные с технологиями работы с данными, робототехники и другими новыми технологиями.

Кроме того, использование технологий работы с данными уже приводит к исчезновению границ между различными направлениями деятельности и исследований в них. Медицинские системы постепенно консолидируют данные из различных узкоспециализированных медицинских систем, формируя фактически «полный профиль пациента», обобщая его данными из отраслевых, национальных и международных реестров. Работа в этом направлении постепенно позволит консолидировать информацию и из других, на первый взгляд, не связанных со здравоохранением систем, таких как данные из различных приложений, собирающих данные об активности и статистику занятий спортом, информацию о питании, качестве сна, показателях кислорода в крови, температуре тела, сердечном ритме, местонахождении и различных других данных. Тем самым будут стерты границы между медициной, спортом и другими науками о жизнедеятельности. Сформируется новый единый массив данных о человеке, анализируя который можно будет не только давать рекомендации, повышающие качество жизни, но и станет возможным избежать появления болезней, предупреждать ухудшение состояния здоровья, повысить качество тренировок и сна и зачастую значительно продлить жизнь человека.

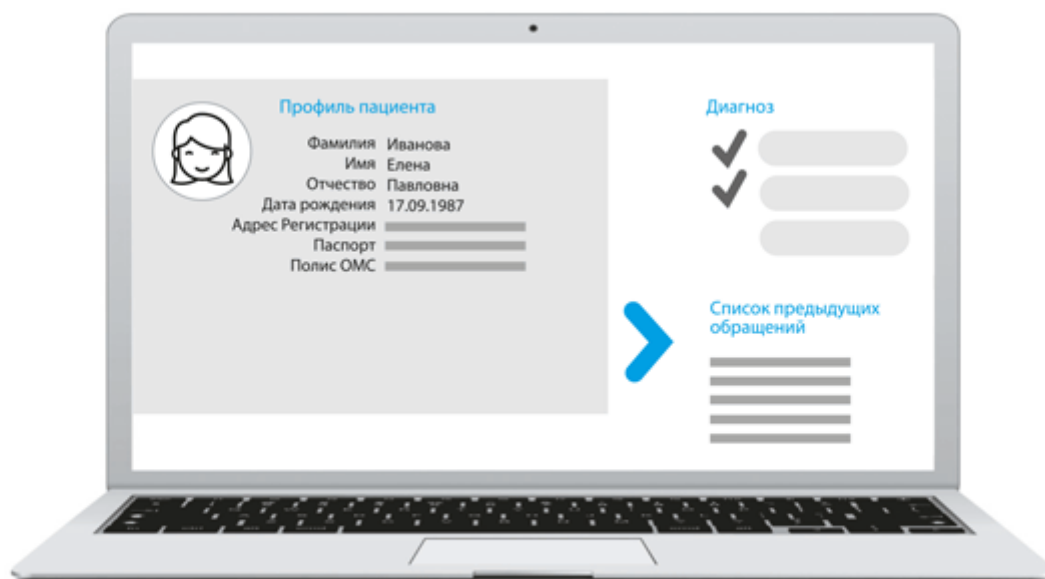


Рис. 18.5.

Профиль пациента

На уровне формирования государственной политики все эти технологии ведут к повышению качества

планирования мер государственной поддержки различных сфер жизни общества, оценки результатов государственных программ и получения оперативной информации по различным срезам показателей здоровья и качества жизни общества, а использование технологий ИИ, как мы уже подробно рассказывали ранее, позволит создать предиктивные модели принятия решений, реализовать систему рейтингов, т. е. объективных оценок результатов работы на разных уровнях управления. В будущем развитие технологий работы с данными приведет к формированию рейтингов пациентов, мотивирующих их вести здоровый образ жизни, заниматься спортом, участвовать в программах вакцинации (что очень актуально в период эпидемии), проходить плановую диспансеризацию для получения различных преференций и более выгодных предложений от медицинских организаций, страховых и транспортных компаний, организаций, работающих в сфере туризма и отдыха, магазинов, банков, спортивных клубов и многих других государственных и частных организаций.

Отдельно стоит отметить, что технологии работы с данными не только существенно помогут докторам улучшить качество работы, но и позволят гораздо более точно ставить диагнозы и назначать лечение, контролировать его течение, своевременно реагировать на изменения здоровья человека и вносить необходимые коррективы. Пациенты же в свою очередь смогут через обратную связь получить полную и объективную информацию о медицинских учреждениях, медицинских работниках, справочные данные о медицинских препаратах, назначенном лечении и т. д. Также цифровые технологии позволят ввести персонифицированную систему рейтингов медицинских работников. Рейтинг будет формироваться на основе многих показателей, включая образование, курсы повышения квалификации, наличие научных степеней, опыт работы и оценок пациентов и профессионального сообщества.

18.2.2. Розничная торговля

Розничная торговля и связанная с ней логистика – сфера, где внедрение технологий работы с данными имеет безграничный потенциал для оптимизации бизнес-процессов и повышения экономической эффективности. Прежде чем мы перейдем к описанию решений и прикладным примерам, хотелось бы остановиться на тех изменениях, которые произошли за последние два года «благодаря» распространению COVID-19. Основные изменения коснулись модели продаж, огромная доля продаж переместилась из офлайн-магазинов в онлайн (приложения и интернет-магазины). Люди вынужденно попробовали массово приобретать товары в электронном виде... И им понравилось! Более того, поначалу существенно вырос размер среднего чека, а магазинам пришлось на ходу перестраивать работу, серьезно умуощняя свои службы доставки. Многие оказались не готовы к возросшим объемам и необходимости сокращения сроков доставки по требованиям покупателей. Некоторые сети задействовали в качестве служб доставки сервисы такси. Эти изменения привели также к цифровизации смежных услуг, в первую очередь оплаты (увеличение использования банковских карт), увеличение участников программ лояльности (карты постоянного клиента, предоставляющие различные преференции и скидки). Все это позволило розничным сетям получить огромный объем информации как о своих клиентах, так и по различным срезам опроса на продукцию своих магазинов.

Благодаря всему этому объем онлайн-продаж вырос настолько, что некоторые стали закрывать свои офлайн-магазины, которые при сопоставимой стоимости продукции существенно менее экономически эффективны, поскольку вынуждены нести постоянные затраты на аренду помещений, оборудования, заработную плату многочисленного персонала, оплачивать коммунальные услуги, обеспечивать запас продукции «на полках» и складе магазина. И это в то время, как онлайн-модель позволяет избежать всех этих затрат и обеспечить общение с покупателем в электронном виде, где стоимость «полок» на порядки ниже, а эффективная организация логистики – несущественна в цене товара. Вместе с тем продажи в электронном виде позволили сформировать поистине гигантский объем данных о своих покупателях, их местонахождении, предпочтениях как по заказываемой продукции, так и по удобному времени доставки, способу оплаты, их персональным данным (информация о семье, размере одежды, возрасте и многое другое). В случае с продуктами питания, бытовой химии и другими средствами первой необходимости на основе анализа спроса на продукцию онлайн-магазины фактически получили возможность предиктивно формировать корзину «часто покупаемых товаров», и оказалось, что люди зачастую покупают одну и ту же продукцию с определенной периодичностью. Все эти закономерности позволили не только повысить качество планирования магазинами закупок продукции, сократив сроки хранения товаров на складе, но и

обеспечить своих покупателей «их» товарами более высокого качества, с длительным сроком хранения и т. п. Все это произошло в том числе за счет исключения из цепочки «поставщик – склад – магазин – покупатель» составляющую «магазин».

Ритейл всегда был одним из пионеров в области применения передовых технологий работы с данными, но в последнее время объемы данных и требования к качеству информации существенно выросли, что привело к необходимости пересмотра не только используемых технологий, но и бизнес-процессов. Прежде всего, следует отметить, что одно из основных решений, пронизывающих все бизнес-процессы розничных сетей, – единый каталог товаров, где содержится актуальная информация о наименовании товаров, их стоимости, различных характеристиках, поставщике, особенностях логистики, хранения и возврата и многие другие параметры. Единый каталог товаров – стержневая информационная система для многих других ИТ-систем розницы, таких как системы закупок, расположения товаров «на полках», их местонахождения, логистики, систем кассового обслуживания, складских систем, систем лояльности покупателей, интернет-магазина и других информационных систем. Цифровизация этих процессов позволяет функционировать магазинам как живому, идеально настроенному организму, способному эффективно изменяться при необходимости.



Рис. 18.6.

Современные покупки совершаются не выходя из дома

Стоит отметить, что в сфере розничных сетей такие перемены стали вопросом выживания, поскольку изменения модели продаж существенно усилили конкуренцию как в части качества предоставляемых услуг, так и в части стоимости товаров для розничных сетей и их покупателей. В будущем представляется, что бизнес-модель работы розничных сетей серьезно изменится, и это коснется не только продуктовых сетей, но и модели продаж практически всех товаров. Продуктовые розничные сети кардинально сократят количество магазинов, оставив только те, где оборот продаж и необходимость покупателей существенны. Также останется востребованным формат гипермаркетов, при этом большая часть покупателей перейдет в формат онлайн-покупок, не утруждая себя посещением магазинов. При этом по статистике большая часть людей покупает с определенной периодичностью примерно одни и те же продукты, к которым привыкли. Если говорить о товарах длительного пользования (одежде, бытовой технике и электронике, автомобилях и других транспортных средствах), то здесь, по всей видимости, останутся большие флагманские салоны производителей и мультибрендовые гипермаркеты, где покупатели смогут воочию увидеть продукцию. Посетив физический магазин, можно ознакомиться с функционалом, выбрать и посмотреть отдельные опции, получить представление о размере и померить одежду, выбрать какую-то продукцию, заказать ее; при этом в будущем достаточно будет заказывать товары через интернет-приложение, при необходимости товар можно предварительно померить при доставке. Все это уже реальность нашего времени, и эти тенденции будут только усиливаться и все прочнее входить в нашу повседневную жизнь. Такой формат

удобен не только покупателям, но и позволяет экономить на содержании многочисленных магазинов и персонала.

Совершенно очевидно, что в разных товарных группах будет выработана своя, отдельная модель работы, но общая тенденция будет развиваться в этом направлении. Эти изменения потребуют совершенно иного подхода к работе информационных систем, еще более жестких требований к качеству данных и функциональных возможностей решений в области работы с данными; нагрузка на эти системы многократно возрастет. В дальнейшем использование искусственного интеллекта позволит существенно повысить эффективность функционирования как бизнес-процессов, так и улучшить работу каждой из подсистем. Но оптимальная работа искусственного интеллекта опирается во многом на платформенные решения по управлению данными, поэтому процесс улучшения данных и инструментария работы с ними – важнейший на всех этапах развития ИТ-ландшафта.

Также стоит отметить, что описанная выше модель потребует также серьезной перестройки процессов логистики и наличия качественного адресного каталога, интегрированного как с системами, в которых указываются данные покупателей/заказчиков, так и с системами, обеспечивающими работу складов и транспорта. Довольно важной составляющей представляется функционирование информационных систем, формирующих различную аналитику, позволяющую оперативно принимать управленческие решения и корректировать ранее принятые решения и настраивать бизнес-процессы.

Как мы неоднократно подчеркивали, наличие огромного массива качественных данных и оптимальных решений для работы с ними станут ключевым конкурентным преимуществом практически во всех сферах деятельности. Кроме изменений, которые происходят на наших глазах, в самое ближайшее время мы станем свидетелями новых направлений бизнеса в сфере розничной торговли. Прежде всего, можно прогнозировать появление компаний, обеспечивающих консолидацию и доставку огромного перечня товаров при отсутствии собственных офлайн-магазинов и складов. Этот формат предполагает крайне эффективную логистику, наличие чрезвычайно удобного пользовательского онлайн-приложения и мощного программного и аппаратного обеспечения. Фактически, речь будет идти о создании огромных цифровых гипермаркетов, работающих полностью в цифровом формате с использованием новейших информационных технологий по работе с данными и логистики. Эффективность подобного бизнеса будет возрастать по мере расширения предлагаемой номенклатуры продукции и географического развития. Первые глобальные компании в этом сегменте уже демонстрируют свои глобальные амбиции, например Amazon. Использование единых платформенных решений, включающих в себя функциональные блоки по каталогизации номенклатуры, закупок, программ лояльности, приложений для покупателей, внутренних и внешних логистических операций, и других приложений, использующих общую платформенную среду, позволяюткратно повысить эффективность операций по сравнению с текущими моделями бизнеса.

18.2.3. Недвижимость

Развитие различных ИТ-технологий, в первую очередь решений по управлению данными, уже оказывает существенное влияние на революционные изменения в различных сферах, связанных с недвижимостью. Уже сегодня появляется все больше разнообразных сервисов и приложений, позволяющих покупать, продавать, арендовать, сдавать в аренду, проверять правовой статус объектов недвижимости и осуществлять различные операции в цифровом виде. Вместе с тем получают широкое распространение и другие электронные сервисы в сфере недвижимости, такие как страхование, нотариальное оформление, кредитование и другие банковские услуги, всевозможные услуги проектирования, проведения строительно-монтажных работ, обслуживания, клининга. Отдельно стоит упомянуть разнообразные государственные сервисы в сфере недвижимости, количество которых также увеличивается с каждым днем. И эта тенденция приобретает все более всеобъемлющий характер, угрожающий перевести все индустрии, связанные с недвижимостью, в цифровой вид.

Электронные сервисы позволят не только достичь небывалого ранее уровня удобства и качества услуг в данной области, но и кардинально изменят рынок услуг, заменив фактически все офлайн-услуги на онлайн-сервисы. В этой новой цифровой парадигме одним из ключевых факторов, обеспечивающих корректную работу сервисов, являются инструменты работы с данными. Качество данных приобретает принципиально важный характер, поскольку основные данные фактически и обеспечивают работоспособность указанных выше сервисов, в то время как объем транзакционных данных будет не

столь существенно влиять на предоставление услуг в сфере недвижимости. По всей видимости, большая часть современных агентств недвижимости, страховых и финансовых организаций, других компаний, работающих в сфере недвижимости, станут цифровыми, формируя различные платформенные решения.

В данной отрасли уже сформировался и будет довольно бурно развиваться сегмент электронных решений для обеспечения комфортного проживания в дальнейшем. Речь идет о различных интерпретациях умного дома – в виде платформ, собирающих различные данные с многочисленных датчиков и управляющих функциональностью разнообразных домашних устройств. В качестве примеров, работающих уже сегодня, это решения Samsung, Miele, Philips, Legrand Netatmo и многих других компаний.

Отдельно необходимо упомянуть возможные синергии от использования интеграции сервисов с различными электронными услугами, предоставляемыми государством. Например, при принятии решения о приобретении или сдачи в аренду недвижимости сразу можно получить расчет предполагаемого налога, а потом и оплачивать его одним кликом, также синергия возможна при страховании недвижимости и ее стоимости и др. Все упомянутые сервисы позволяют собрать такой объем данных, который позволит проводить не только рыночные исследования в области недвижимости принципиально другого качества, но и создаст основу для услуг, предоставляемых банками, страховыми и другими компаниями следующего поколения, которые сейчас немыслимы и невозможны, но сформируют новые сегменты бизнеса. Все это позволит оказывать частным и государственным организациям качественные и прозрачные услуги, позволяющие также обеспечить существенную экономию и избежать возможного мошенничества.

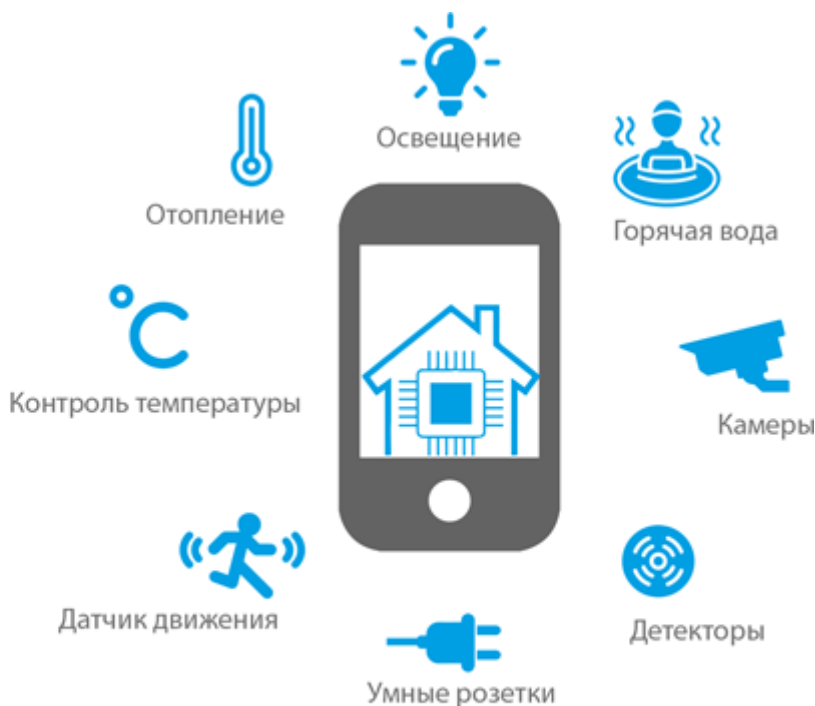


Рис. 18.7.

Современный умный дом имеет множество функций

Помимо этого, цифровизация сферы недвижимости откроет совершенно новые возможности для различных коммунальных компаний и структур, специализирующихся на обеспечении безопасности, логистических услугах и многих других. Отметим, что все цифровые сервисы также востребованы и в части коммерческой недвижимости. Благодаря этому корпоративный сегмент станет также частью электронных сервисов и их клиентом. Кроме того, данные этих сервисов будут востребованы государственными и коммерческими организациями, поскольку позволят формировать аналитическую отчетность на совершенно ином качественном уровне по сравнению с текущим положением вещей.

18.2.4. Транспорт

Индустрия транспорта – одна из основных сфер применения информационных технологий и решений по работе с данными. Практически повсеместно уже используются всевозможные электронные реестры: транспортные средства, водители, пассажиры, различные комплектующие, разрешения и удостоверения, реестры прохождения плановых ремонтов и технического обслуживания, различные расписания и другие критически важные данные. Дальнейшее развитие технологий управления данными потребует полной интеграции всех систем, содержащих критически важные данные, посредством специализированных платформ управления данными. Именно такие платформы занимают важнейшую нишу интеграции сложнейших информационных систем и обеспечивают корректную работу различных функциональных систем практически всех крупных государственных и коммерческих организаций, в том числе в сфере транспорта.

Транспортная отрасль характеризуется высокой степенью качества как критически важных, так и транзакционных данных, поскольку эта информация не только обеспечивает бизнес-процессы организаций, но и существенно влияет на безопасность и работоспособность этих организаций.

Здесь важно отметить, что транспортная индустрия сейчас очень сильно трансформируется. Мы живем в эпоху революционных технологических изменений в сфере транспорта. В самое ближайшее время мы станем свидетелями кардинальных изменений в отрасли, когда средства транспорта станут цифровыми во всех смыслах. В качестве примера приведем автомобильный транспорт. Все больше производителей анонсируют свои планы по полному отказу от производства автотранспорта на двигателях внутреннего сгорания в пользу электрической тяги. По некоторым данным, к 2035 году практически все производимые в мире автомобили будут электрическими. Одно из основных направлений работы в области исследований и разработок автоконцернов сейчас – это работы в области более эффективных батарей, позволяющих автомобилям преодолевать большие расстояния без необходимости подзарядки или существенного сокращения времени зарядки. Параллельно с этим по всему миру создается инфраструктура для электромобилей, в первую очередь пункты зарядки автомобилей. Технологии в данном направлении развиваются очень быстро, и совсем скоро привычные нам автомобили станут редкостью и со временем исчезнут совсем.

Кроме того, по результатам многочисленных исследований в подавляющем большинстве автомобилей перевозит всего одного человека, в то время как большинство современных машин рассчитано на 4–5 человек. Именно поэтому автомобили в будущем станут гораздо компактнее, и в большинстве своем будут рассчитаны на одного-двух человек. Также одним из перспективных направлений исследований в этой области являются технологии беспилотного управления (например, эту технологию развивают Tesla и Google). Именно развитие беспилотного транспорта становится ключевым фактором ускоренного развития технологий работы с данными в транспортной отрасли. Агрегация данных транспортной сферы позволит обеспечить формирование целого сегмента новых инновационных сервисов. Как было отмечено ранее, помимо различных реестров основных данных, новое поколение транспортных средств формирует огромный объем транзакционных данных, позволяющих не только обеспечивать режим движения транспортных средств, но и получать целый массив телематической информации, данных о техническом состоянии транспорта и множество другой информации вплоть до сведений о самочувствии пассажиров. Вместе с тем широкое использование технологий работы с данными позволит обеспечить эффективные механизмы администрирования отрасли со стороны государства, отслеживая всевозможные нарушения лицензионной политики, технического состояния транспортных средств, самочувствия водителей и многих других параметров.



Рис. 18.8.

Современная транспортная сеть — больше чем просто перевозки

Отдельно стоит упомянуть и грядущие изменения в транспортной инфраструктуре, а именно интеллектуальные дорожные знаки, разметка дорог «активной» краской, различные технологии освещения и обеспечения безопасности. Кроме того, инновации в транспорте – и в первую очередь решения по управлению данными – приведут к существенному улучшению экологических норм и требований, обеспечивая снижение всевозможных выбросов и другого негативного влияния на окружающую среду. Также цифровые технологии формируют совершенно новую парадигму логистических решений, делая оптимальными маршруты для всех видов транспорта, реагируя и оптимизируя движение в режиме реального времени.

18.2.5. Сельское хозяйство

Сельское хозяйство – одна из наиболее перспективных сфер применения платформенных решений в области управления данными. Уровень развития инновационных решений в этой области уже сейчас позволяет обеспечить полномасштабную цифровизацию основных бизнес-процессов отрасли. Уже сейчас имеются технологии работы с данными, обеспечивающие ведение реестра недвижимости, включая сельскохозяйственные земельные участки в цифровом виде с довольно богатым атрибутивным составом, описывающим практически все характеристики этих земельных участков: размер, географическое положение, вид разрешенного использования, наличие зданий и сооружений, данные владельца, договоры и характеристики аренды, показатели почвы и урожайность. В связи с тем, что основой всех бизнес-процессов в сельском хозяйстве является земля, то и все данные, характеризующие работу сельскохозяйственных предприятий, также формируются на основе базовой информации о земле.

Объем данных в сельском хозяйстве поистине огромен. Современные отраслевые информационные системы позволяют собирать и консолидировать информацию об истории использования сельхозугодий, урожайности, химическом составе почв, метеорологии, использовании различных удобрений и средств механизации, многочисленных показателей животноводческого комплекса в различных разрезах. Кроме того, консолидация разнообразных данных, их качество и полнота позволяют перевести многие сервисы в цифровой формат, а также обеспечить условия для формирования новых инновационных электронных систем, таких как возможность полноценной оценки сельхозпроизводителей со стороны банковских и страховых организаций, поставщиков различной продукции, лизинговых компаний. Благодаря технологиям управления данными появляется возможность обеспечить актуальной информацией органы государственного администрирования и сертификации.

Сервисы на основе технологий управления данными приведут также к совершенно новому качеству приложений для потребителей как розничных сетей, так и покупателей оптовой продукции, обеспечивая их всей полнотой информации и ее характеристиках предлагаемых товаров. Одно из ключевых систем работы с данными – внедрение на высоком уровне Единого каталога товаров, работ и услуг (ЕКТРУ), аналога международного GS1. Благодаря этому обеспечивается единая информационная среда на всех уровнях производства и потребления продукции и сервисов, а администрирование отрасли становится эффективным и прозрачным. Также внедрение ЕКТРУ позволит существенно повысить качество функционирования систем цифровой маркировки товаров, логистики, а также налогового и таможенного администрирования. Еще одним направлением исследования технологий работы с данными в сельском хозяйстве являются многочисленные системы сбора транзакционных данных с различных датчиков и других телематических устройств, позволяющих обеспечивать продукции, животноводческого комплекса, санитарного и ветеринарного контроля. Для многих сельскохозяйственных предприятий использование подобного инструментария станет эффективным решением повышения качества продукции, снижения издержек и значительного улучшения экономических показателей.

18.2.6. Жилищно-коммунальное хозяйство

Не будет преувеличением сказать, что каждый из нас практически каждый день сталкивается с услугами жилищно-коммунального хозяйства (ЖКХ), и эти услуги во многом определяют качество жизни жителей современного мегаполиса. Мы привыкли к тому, что предоставление услуг холодного и горячего водоснабжения, электроэнергии, отопления, вывоза мусора, содержания подъездов и общественных мест, обслуживания подъемных механизмов/лифтового хозяйства, освещения общественных мест, канализации является неотъемлемой частью нашей жизни. При этом ЖКХ – это огромная индустрия, и эффективность ее деятельности в современных условиях во многом зависит от используемых технологических решений. Они могут быть индустриальными (и тогда мы видим современные электрические устройства, различные механизмы обеспечения холодного и горячего водоснабжения, современные трубы с низким коэффициентом теплопотерь, использование новых лифтовых механизмов, энергоэффективных технологий освещения) и связанными с технологиями работы с данными и различными решениями диспетчеризации.

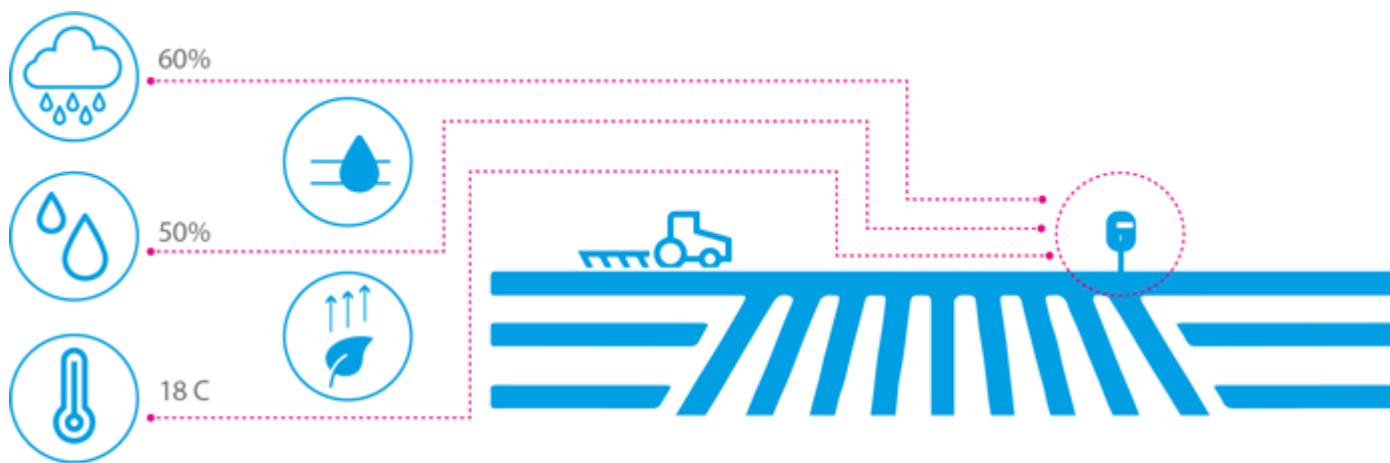


Рис. 18.9.

Современные датчики следят за климатическими условиями на полях

Уже сейчас предприятия, работающие в сфере ЖКХ, активно используют разнообразные автоматизированные системы управления, позволяющие обеспечить эффективные технологии диспетчеризации и сбора различных данных и показателей с приборов учета и управления. Здесь традиционно все данные можно разделить на основные данные абонентов (фамилия, имя, отчество), объект недвижимости (со множеством своих атрибутов), устройства учета (со своими уникальными номерами, данными сертификации, поверки, производителями и поставщиками, местами установки,

обслуживающими организациями, наименованиями информационными системами сбора и биллинга и другой информации) и транзакционные данные, получаемые со всех приборов учета (различные показатели устройств, в том числе в режиме реального времени, информацию об их работоспособности и др.). Все это позволяет обеспечить абонентов качественными и экономически обоснованными услугами, а также интегрировать подсистемы сбора данных с корпоративными системами управления ресурсами посредством единого слоя хранения и управления данными.



Рис. 18.10.

Будущее ЖКХ в действии

Накопление и анализ огромного количества данных позволит крупным отраслевым организациям обеспечить эффективное использование оборудования. Так, консолидация данных на уровне региональных поставщиков электроэнергии позволит более эффективно и качественно использовать имеющиеся ресурсы, такие как выравнивание графиков энергопотребления, повысить надежность и качество энергоснабжения, а использование технологий искусственного интеллекта – перевести в автоматический режим управление генерацией и распределения электроэнергии на основе данных телеметрии и телемеханики, статистических качественных данных в режиме реального времени, сделав первый шаг в переводе основных бизнес-процессов и управление энергосетями в режим автопилотного управления.

Вместе с тем использование решений для работы с данными также формирует основу для внедрения корпоративных сервисов, позволяющих обеспечить эффективные решения для закупок товаров и услуг, управления персоналом, включая управление мобильными бригадами, разработку различных интерпретаций «Личного кабинета абонентов» и других сервисов технического ремонта и обслуживания оборудования. Словом, всех инноваций, базисом для которых являются платформы управления данными и сами данные, с которыми в постоянном режиме проводятся работы по обеспечению их полноты, качества, непротиворечивости, интероперабельности.

18.2.7. Природопользование

Сфера природопользования с каждым годом становится все более значимой в экономических отношениях государства и бизнеса. Современные технологии уже сегодня позволяют минимизировать влияние вредных производств на экосистему, а зачастую использовать природные ресурсы для повышения не только эффективности производств, но и обеспечить оптимальность их размещения, логистики, прозрачности и экологичности выпускаемой продукции, а также формирование новых, инновационных направлений деятельности. В этой сфере востребован огромный спектр различных технологий, имеющих как узкоотраслевое применение, так и универсальное направление, используемое в самых разных

кластерах и сегментах, относящихся к сфере природопользования.

Одним из таких универсальных технологических направлений являются всевозможные ИТ-решения и технологии работы с данными. С учетом специфики отрасли все большую популярность приобретают различные централизованные информационные системы уровня как небольших предприятий, так и крупных корпораций и даже государственных органов власти. Как и во многих других отраслях экономики, основу большинства современных ИТ-систем формируют массивы основных и транзакционных данных. В части основных данных необходимо отметить, что сбор и формирование этого слоя информации должны осуществляться централизованно и единообразно. Это может включать в себя все многообразие нормативно-справочной информации, а также данные, необходимые для работы служб материально-технического обеспечения, финансово-экономических показателей, сервисов работы с поставщиками и покупателями, служб обеспечения качества продукции и ее сертификации и лицензирования, а также бизнес-процессов, обеспечивающих учет и прозрачность происхождения продукции.

Здесь в качестве примера можно привести разнообразные информационные системы контроля бизнес-процессов выпуска и заготовки продукции на уровне государства, позволяющие повышать прозрачность в таких сферах, как лесозаготовка, лесопереработка, заповедные территории, всевозможные виды охоты, вылов и переработка морских и речных ресурсов, сбор и переработка мусора и многие другие направления деятельности. Практически во всех информационных системах подобного класса качество основных данных – фундамент корректной работы этих решений.

В этой сфере поистине огромный объем информации, позволяющий обеспечить построение крупных информационных систем, которые работают с качественными основными данными и огромным массивом транзакционных данных. Это данные, которые в режиме практически реального времени генерируются различными датчиками, следящими за состоянием экологической сферы, качества воздуха, химического состава водоемов и почв, местоположением представителей животного мира, показателей очистных сооружений, изменении температурного режима почв, воздуха и воды. Помимо этого, большой объем данных содержится в различных информационных системах органов государственной власти и компаний, обеспечивающих администрирование со стороны государства и управления бизнес-процессами коммерческих организаций. Весь этот массив данных позволяет не только существенно повысить качество администрирования и управления, но и ведет к кардинальному изменению всей сферы природопользования.

18.2.8. Промышленное оборудование

Ни для кого не секрет, что каждое следующее поколение промышленного оборудования становится все более сложным. Используется все больше технологических решений, основанных на данных, например, количество и точность различных встроенных датчиков, измеряющих разнообразные показатели и режимы работы. В свою очередь усложняется и сама продукция. Все это привело к тому, что оборудование уже позволяет обеспечить беспилотную работу, а многочисленные датчики и решения искусственного интеллекта формируют промышленные кластеры производства товаров и услуг, работающих в режимах, близких к автономным. Отдельно стоит упомянуть, что активное использование робототехники и решений в области управления данными ведет не только к снижению человеческого фактора в промышленности (здесь речь может идти о повышении качества и точности производства, существенном увеличении производства, повышении безопасности, кардинальном снижении себестоимости продукции, новых технологиях обеспечения гарантии технического сопровождения и продаж), но и позволяет производственным компаниям менять подходы к развитию всего бизнеса.

Так, при рассмотрении вопроса о расширении производства все больше внимания уделяется стоимости электроэнергии, логистике, возможности привлечения высококвалифицированных кадров. Все меньшую роль играет стоимость низкоквалифицированного персонала, который постепенно заменяется робототехникой, станками с ЧПУ, различными информационными системами, которые обеспечивают работу с заданными параметрами качества, точности и объема, что позволяет организовать производство в круглосуточном режиме с входным и выходным контролем выпускаемых изделий без влияния человеческого фактора. Эти преимущества использования высокотехнологичных решений в производстве уже заставляют крупные международные корпорации задумываться о географии развития своих производственных площадок без привязки к стоимости низко и среднеквалифицированного персонала,

который в самое ближайшее время будет заменен различными технологическими решениями, большая часть которых имеет возможность автопилотирования и удаленного управления. В этой связи при развитии или создании производств на первый план выйдут условия обеспечения инфраструктурой и различные условия государственного регулирования: стоимость подключения к инфраструктуре, дороги, налогообложение, стоимость и условия кредитования, различные страховые и лизинговые механизмы и другие условия, являющиеся основой для принятия решений о стоимости и возможности создания и развития производства.

Хотелось бы отдельно обратить внимание читателя на изменения, происходящие в области продаж и технического сопровождения сложного технологического оборудования и продукции. Сейчас все чаще стоимость продукции указывается не одной суммой, которую необходимо единовременно оплатить, а в виде стоимости, где условия оплаты привязаны к параметрам использования продукции, что стало возможным благодаря развитию технологий создания датчиков, решений в области передачи и обработки данных. В разделе 18.1.3, посвященном интернету вещей, мы подробно рассказывали о компании General Electric, уже длительное время продающую турбины для самолетов и систем энергетики по часам налета/использования, что позволяет эффективно использовать финансовые ресурсы. В будущем подобная модель, предполагающая продажу сложного и дорогого оборудования, будет распространяться на подавляющее большинство сделок, оплата по которым будет привязана к единицам измерения оборудования (часы налета, километры пробега, качество выпущенной продукции, время работы оборудования в различных режимах и другие). Все это стало возможным благодаря широкому распространению решений по работе с данными. Кроме того, подобный формат позволит производителю удаленно обеспечить сбор статистики по всем необходимым показателям для корректной стоимостной оценки, а также для обеспечения необходимого уровня качества работы оборудования, обновления программного обеспечения, оценки износа оборудования и его компонентов и многих других параметров. Производитель в подобном формате сможет собрать огромный массив данных со всей продукции, работающей у его заказчиков. Такой анализ поможет получить объективную и полную информацию, необходимую для принятия управленческих решений, а также для формирования новых направлений развития бизнеса.



Рис. 18.11.

Будущее промышленного производства

18.2.9. Логистика

В последние годы под влиянием бурного развития различных инновационных технологий, а также с появлением новых вызовов для всего человечества, таких как новые заболевания, с которыми ранее люди не сталкивались, с высоким показателем смертности и легкостью инфицирования новых заболевших,

произошли поистине революционные изменения в области логистики и технологий доставки товаров и услуг. COVID перевернул традиционное представление людей о привычной размерности жизни, в которой многие обычные ранее вещи, такие как посещение магазинов, различных организаций, предоставляющих разнообразные услуги, аптеки, больницы, культурные массовые мероприятия, стали небезопасным с точки зрения заражения вирусом.

На национальном и международном уровнях достаточно оперативно были введены различные ограничения, чтобы не допустить существенного роста числа заболевших. В первую очередь это меры, препятствующие личному общению и контактированию с людьми. COVID стал глобальным и угрожающим фактором, требующим изменений большинства индустрий, обеспечивающих многие сферы жизнедеятельности людей на всей планете, что послужило мощным толчком для ускорения, переосмысления и изменения форматов внедрения разнообразных технологических решений, исключающих личное общение между людьми. Все это разорвало вековые традиции и привычки межличностных отношений. Еще несколько лет назад широкое распространение удаленной работы и ограничение личного общения казалось чем-то невероятным, немыслимым, поскольку личное общение – фундамент развития культуры, бизнеса, науки, политики и других сфер жизнедеятельности. На основе личного общения стало возможно появление многих изобретений, и современные мировые мегаполисы обязаны этим коммуникациям. Борис Джонсон в своей книге «Лондон по Джонсону: О людях, которые сделали город, который сделал мир» приводит большое число примеров важности и результативности подобного общения, обобщая это выводом о том, что «люди хотят встречаться с другими людьми лично», и описывает большое количество изобретений и открытий, создавших целые индустрии и современные города мира, столицы культурной, деловой и научной жизни человечества⁵⁵⁴.

Столкнувшись с COVID, человечеству пришлось перестраиваться практически во всех сферах жизнедеятельности. В первую очередь это затронуло те направления жизни, которые позволяют обеспечить формат удаленного общения, жизни и работы: различные технологические сервисы, в основном построенные на информационных технологиях, новые решения в области работы с данными в широком смысле слова; решения в сфере транспорта, логистики и предоставления различных услуг доставки; различные сервисы и возможности приобретения товаров и услуг в удаленном режиме, включая сервисы развлечений (всевозможные онлайн-кинотеатры, концерты, игры и многое другое), бизнес-приложения для онлайн-конференций, совещаний, вебинаров, совместной работы в производственных системах, онлайн-голосования и всевозможные опросы и огромное количество других решений. Те изменения, которые мы сейчас видим, – это лишь первый шаг к гораздо более масштабным изменениям, которые ожидают человечество как в отдельных направлениях деятельности, так и процессе появления новых индустрий и технологий на основе синергии отдельных современных сфер экономической, научно-исследовательской, культурной и других направлений человеческой мысли.

⁵⁵⁴ Джонсон Б. Лондон по Джонсону: О людях, которые сделали город, который сделал мир / Борис Джонсон; пер. с англ. О. Алексанян. – М.: Азбука-Аттикус, 2017.



Рис. 18.12.

«Шестеренки» современной логистики

Одно из предположений о наиболее вероятных переменах связано с появлением новых видов коммуникаций для быстрой и надежной доставки и отправления всевозможных товаров. Уже существуют тысячи компаний, занимающихся исследованиями и разработками в области создания летательных аппаратов для этих целей. Они используют в том числе и технологии искусственного интеллекта, обеспечивая беспилотный режим управления большим массивом одновременно задействованных аппаратов, фактически сделав огромный прорыв в сфере логистики. Можно предположить, что лидеры в развитии подобных технологий, такие как Amazon, еще более упрочат свои лидерские позиции и станут активно вытеснять компании, использующие устаревшие технологии. С большой долей вероятности через несколько лет исчезнет профессия курьера и множество сопутствующих видов бизнеса. Возможно, эта сфера пройдет тот же эволюционный путь что и телекоммуникационная индустрия, а ключевым показателем эффективности и роста бизнеса станет так называемая последняя миля (т. е. заключительный этап доставки, в дистрибуции – это доставка до розничных торговых точек а в электронной коммерции – доставка из сортировочного центра до конечного адресата), и компании, имеющие лучшие технологии ее обеспечения, станут лидерами в области логистики, вытеснив своих неконкурентоспособных коллег с рынка.

В ближайшее время технологии поставки «по воздуху» вряд ли получат широкое распространение в крупных мегаполисах по целому ряду вполне объективных причин: сложность диспетчеризации; огромное количество операций, риск получения травм при авариях; предельные емкости батарей, ограничивающих время автономной работы; всевозможные регламенты безопасности как городов, так и их жителей и многие другие факторы, ограничивающие использование летательных средств доставки в городах.

Гораздо более вероятно появление совершенно новых ниш для использования инновационных технологий в сфере логистики для «последней мили». Одним из возможных и крайне востребованных проектов вот-вот станет решение вакуумной почты для крупных мегаполисов. С учетом развития технологий вакуумных поездов – высокоскоростного вида транспорта, движущегося в сильно разреженном воздухе с помощью магнитной левитации, стоит отдельно упомянуть проект Hyperloop, разрабатываемый под руководством Илона Маска, для доставки пассажиров на специальных поездах на скорости более 1200 км/ч. Возможно, что подобные решения могут быть использованы для создания вакуумной почты, учитывая, что вся техническая документация открыта и доступна для изучения. Эти технологии могут привести к поистине революционным изменениям в области доставки и отправления товаров. Оснащение жилых, офисных и торговых помещений вакуумными почтовыми ящиками, связанными между собой разветвленной сетью вакуумного трубопровода с огромным количеством разнообразных датчиков,

позволит создать совершенно новую индустрию логистики товаров на основе новейших технологий, воплотив в жизнь то, что ранее мы видели только в фантастических фильмах. Это еще один этап технологического развития человечества в его прикладном, обеспечивающим комфорт смысле, наряду с централизованной канализацией, горячим и холодным водоснабжением, электрическими и телекоммуникационными сетями, без чего уже сложно представить жизнь современного человека.

Эта вакуумная почта будет «заходить» в каждую квартиру, каждый дом, офис, магазин, склад, аптеку, больницу, министерство и ведомство, что позволит обеспечить фактически мгновенную доставку любого товара или документа. Словом, практически всего того, ради чего мы посещаем магазины, офисы, различные учреждения и бытовые службы, либо того, что заказываем в интернет-магазинах. Подобная система, по всей видимости, в будущем будет пронизывать все современные мегаполисы, а при введении единых стандартов и ее функционирования, вакуумная почта позволит обеспечить междугородний и международный оборот доставки. Как и большинство инновационных технологий, после широкого распространения вакуумная почта изменит либо заменит собой все связанные с доставкой товаров индустрии. Прежде всего это приведет к исчезновению служб почтовой доставки в классическом понимании этого слова, сферы курьерской и экспресс-доставки документов, продуктов питания, готовой пищи, лекарств, крупногабаритных товаров. Ощутимый эффект почувствуют на себе также и магазины розничной торговли, включая розничные сети. Все это кардинально изменит бизнес-процессы в большинстве индустрий, обеспечивающих жизнь современного человека.

С развитием вакуумной почты изменение претерпит и вся индустрия производства упаковки товаров, которая вынуждена будет перестроиться под требования и стандарты «почты нового поколения». Существенно изменится работа компаний, предоставляющих различные услуги: от ресторанного «удаленного» обслуживания (доставка готовых блюд к заданному времени) и бытовых услуг (например, получение и доставка вещей для чистки) до доставки всевозможных товаров. Это очень похоже на работу телекоммуникационных сетей, но отличие в том, что вместо передачи пакетов данных в вакуумной почте используется передача товаров. Помимо служб доставки и всевозможных магазинов розничной торговли, в сегодняшнем нашем представлении, пострадают также производители бытовой техники для дома (холодильники, стиральные и сушильные машины, СВЧ-печи и другие приборы, обеспечивающие нас горячим питанием, возможностью хранения продуктов, уходом за одеждой и многим другим, позволяющие нам «автономно» существовать). Это же актуально и для офисов и различных учреждений. Фактически везде в будущем нас будут окружать вакуумные почтовые ящики, позволяющие ими пользоваться при помощи различных приложений для компьютера и смартфона.

Подобные инновации в области доставки товаров и услуг потребуют самых передовых решений в технологиях работы с данными. Как и в большинстве масштабных информационных систем индустриального уровня, основу и качество их функционирования обеспечивают разнородные данные, их качество и объем, а также технологии их обработки и управления. Для корректной работы подобной информационной системы требуется поистине огромный массив данных и обеспечение их качества. Это информация о владельцах вакуумных почтовых ящиков (фамилия, имя, отчество, удостоверяющие личности документы, фактический адрес, контактные данные, платежные реквизиты и много другое, описывающие как владельца ящика, так и членов его семьи, коллег по работе и другую информацию), перечень всех абонентов инновационной почты, магазины с реестрами номенклатуры товаров и их подробным описанием, соответствующих единому каталогу товаров, работ и услуг всей почтовой системы, информации о самом магазине или организации, а также других организациях и учреждениях. Развитие подобных решений приведет к взрывному росту объема данных, и сейчас даже сложно прогнозировать, как это отразится на технологиях управления данными и на различных сопутствующих ИТ-направлениях. Эффект от подобного масштаба инноваций будет всеобъемлющий.

Пример, описанный выше, говорит о том, что на основе решений в области управления данными открываются огромные возможности для предпринимательских талантов, развития существующих идей и компаний, позволяющих совершать фактические проекты в различных направлениях жизни людей, способных изменить жизнь человечества совершенно невообразимым образом. Стоит отметить, что все чаще можно услышать, что информационные технологии, в частности программное обеспечение, играют доминирующую роль в создании и развитии совершенно новых решений во многих сферах, далеко выходящих за рамки традиционных областей, относящихся к высоким технологиям. Здесь можно

процитировать книгу «Блиц-масштабирование: как создать крупный бизнес со скоростью света»: «Индустрии, которые сосредоточены на физических продуктах (атомах), объединяются с программным обеспечением (биты). Tesla создает машины (атомы), но обновления ПО (битов) может сразу улучшить разгон этих машин и добавить автопилот»⁵⁵⁵.

Литература к главе 18

• Айзексон У. Инноваторы. Как несколько гениев, хакеров и гиков совершили цифровую революцию / Уолтер Айзексон; пер. с англ. И. Кагановой, Т. Лисовской, О. Храмцовой. – М.: АСТ, 2016.

Заключение

В продолжение темы отраслевых решений на основе недавно сформировавшейся индустрии данных (как мы писали ранее, это уже не часть информационных технологий, а вполне сформировавшееся отдельное высокотехнологическое направление), отметим, что специалисты по управлению данными часто сталкиваются со следующим заблуждением. «Пусть мы пока плохо понимаем, как у нас все работает, какие данные мы собираем и храним, как построены наши бизнес-процессы, на основе каких данных мы принимаем решения, зато сейчас мы реализуем проект внедрения искусственного интеллекта/машинного обучения и все эти лучшие практики позволят нам стать лидерами рынка!»

Подобный неверный подход, увы, довольно часто проповедуется руководителями различных уровней и различных компаний и ведомств. И заканчивается все это печально: через некоторое время организация приходит к неутешительному выводу о невозможности достичь первоначальных целей проектов. Отдельно стоит упомянуть, что прилагаемые усилия по обеспечению качества данных иногда вместо повышения эффективности бизнеса, лояльности клиентов и отдачи от современных технологий, наоборот, приводят к оттоку клиентов, снижению среднего чека, невозможности повышения операционных показателей компаний, а для различных государственных ведомств – к увеличению критики отдельных цифровых сервисов, нежеланию граждан ими пользоваться и необходимости отвлекать сотрудников для постоянной «ручной» работы там, где предполагалось существенное снижение их участия за счет цифровизации. Все дело в том, что внедрение указанных выше технологий, базирующихся на эффективном использовании данных, невозможно без прохождения предварительных этапов их планирования, сбора, хранения и внедрения инструментария для обработки.

В нашей книге мы попытались показать, что ценность информационных ресурсов организации зависит от совокупности многих факторов, включая лидерство и приверженность руководителей принципам грамотного управления данными.

Подведем некоторые важные итоги.

1. Чтобы достойно противостоять деструктивным воздействиям, характерным для четвертой промышленной революции, организации должны извлекать максимальную ценность из имеющихся у них информационных ресурсов (превратить их из «мусора» в «нефть»).
2. Одним из основных препятствий при извлечении ценности является разрыв между данными и информацией в цикле преобразований «данные – информация – знания».
3. Необходимый шаг в преодолении указанного разрыва – изменение отношения организаций к своим данным. Они должны начать рассматривать их как важнейший актив.
4. Выработка правильного отношения к данным связана с преодолением организацией ряда барьеров, включая осуществление перехода на дата-центричное мышление.
5. Лидерство и модели поведения в части управления данными как активом должны исходить от высшего руководства организации.
6. Ценность информационных активов организации устойчиво повышается, только когда

⁵⁵⁵ Хоффман Р., Йе К. Блиц-масштабирование: как создать крупный бизнес со скоростью света. / Рид Хоффман, Крис Йе; предисл. Билла Гейтса; пер. с англ. Юлии Репях. – М.: Альпина Паблишер, 2019.

разрабатывается и внедряется всеобъемлющая корпоративная логистика цепочек поставок данных, поддерживающих цепочки их ценности.

7. В настоящее время наиболее полным и универсальным методическим документом в области управления данными является руководство DAMA-DMBOK2.

8. Рамочную структуру функций управления данными DMBOK2 (в привязке к их жизненному циклу) можно рассматривать как модель управления цепочками поставок данных.

9. Предприятия и учреждения должны сформировать у себя организационные системы руководства и управления данными, отвечающие за эффективную реализацию модели управления цепочками поставок данных.

10. Ключевые первоочередные шаги организации по реализации модели управления цепочками поставок данных:

- оценка текущего состояния;
- использование результатов оценки для планирования улучшений и формирование «дорожной карты»;
- инициирование программы управления организационными изменениями, поддерживающей выполнение дорожной карты.

Все происходящее в области технологий работы с данными не только формирует на наших глазах новую индустрию со всеми необходимыми атрибутами (собственные образовательные программы, различные специальности, новые методы и алгоритмы работы с учетом отраслевой специфики), но и фактически определяет масштаб нарастающих перемен в деятельности большинства организаций и отдельных сотрудников. Это говорит о том, что в будущем управление данными станет одной из ключевых и наиболее востребованных компетенций практически во всех сферах человеческой жизни.

Коллектив авторов книги «Ценность ваших данных» не останавливается на достигнутом. Нам хочется сказать еще многое. И пока эта книга отправляется на верстку, мы уже начинаем работать над новой. Не оставим и эту – ее ждут новые редакции и дополнения. Спасибо вам за внимание к нашему труду!