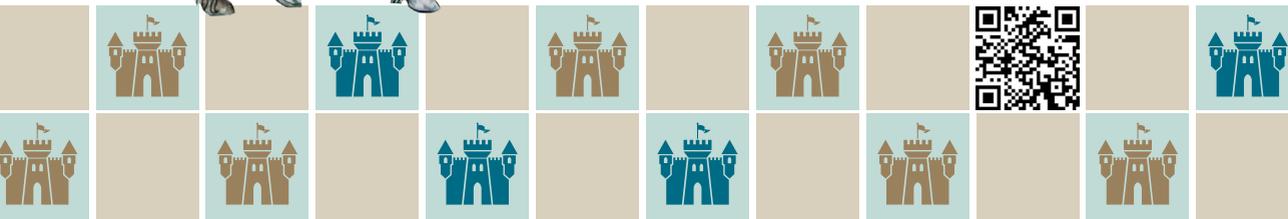


УПРАВЛЕНИЕ ДАННЫМИ В ГОССЕКТОРЕ

НАВИГАТОР ДЛЯ НАЧИНАЮЩИХ



Москва, 2022



УДК 519.25:311

ББК 60.6

У67

Управление данными в госсекторе. Навигатор для начинающих /
У67 под ред. О. М. Гиацинтова, В. А. Сазонова, М. С. Шклярук. – М. :
РАНХиГС, 2022. – 198 стр.

Начиная работу с данными в государственных организациях, руководители цифровой трансформации, команды по работе с данными и прочие госслужащие сталкиваются с рядом проблем: им может не хватать теоретических знаний, опыта работы именно с госданными и информационными системами, навыков аудита имеющейся инфраструктуры. Навигатор знакомит с теоретическими основами управления данными и предлагает конкретные рекомендации, позволяющие наладить эту работу на практике.

Пособие предназначено для госслужащих, вовлеченных в работу с данными, а также для всех интересующихся темой цифровой трансформации госсектора. Оно входит в комплект учебных материалов программы профессиональной переподготовки «Руководитель цифровой трансформации» Центра подготовки руководителей и команд цифровой трансформации.

УДК 519.25:311

ББК 60.6

ISBN 978-5-6044763-0-7

© Российская академия народного хозяйства
и государственной службы
при Президенте Российской Федерации, 2022

АВТОРЫ



Абдрахманов Руслан Ильнурович,
заведующий сектором управления данными, руководитель
портфеля проектов ГКУ «Центр цифровой трансформации
Республики Татарстан»



Анисимова Ирина Александровна,
эксперт федерального центра компетенций ФГБУ
«ВНИИ труда» Минтруда РФ, эксперт Центра подготовки
руководителей и команд цифровой трансформации ВШГУ
РАНХиГС



Архипов Алексей Владимирович,
редактор направления исследований и аналитики
Центра подготовки руководителей и команд цифровой
трансформации ВШГУ РАНХиГС



Афанасьева Лена Николаевна,
канд. мед. наук, министр здравоохранения
Республики Саха (Якутия)



Бойцова Елена Александровна,
директор департамента финансового аудита
Счетной палаты РФ



Виниченко Олег Алексеевич,
исполнительный директор, дирекция «Цифровые решения
для государственного сектора» ПАО «Сбербанк»



Гаркуша Наталья Сергеевна,

д-р пед. наук, руководитель Центра оценки и кадрового развития специалистов в области цифровой трансформации Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС



Гейн Яна Эдгаровна,

дата-аналитик Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС



Гиацинтов Олег Михайлович,

технический директор компании DIS Group



Давыдов Денис Геннадьевич,

канд. психол. наук, доцент кафедры психологии и педагогики Открытого университета экономики, управления и права



Димитров Илья Димитров,

президент ГК Seldon, общественный омбудсмен по вопросам развития цифровой экономики, исполнительный директор НКО «Ассоциация электронных торговых площадок» (АЭТП)



Добровольский Иван Маркович,

руководитель программ ПАО «Газпром нефть»



Жданова Ольга Сергеевна,

начальник отдела цифровой трансформации и государственных сервисов управления технологий цифрового государства министерства цифрового развития и связи Новосибирской области



Ильященко Михаил Александрович,

главный советник отдела управления данными департамента цифровой трансформации Счетной палаты РФ



Кашко Дмитрий Анатольевич,
начальник отдела в департаменте управления данными
ФГБУ НИИ «Восход»



Кзылходжаева Дина Алиевна,
руководитель юридической группы АНО «Центр
перспективных управленческих решений»



Киселева Ксения Львовна,
канд. филол. наук, главный редактор Центра подготовки
руководителей и команд цифровой трансформации
ВШГУ РАНХиГС, ст. науч. сотр. Института русского языка
им. В. В. Виноградова РАН



Константинов Александр Сергеевич,
директор по маркетингу компании «Юнидата»



Коршунова Светлана Вячеславовна,
заместитель руководителя направления исследований
и аналитики Центра подготовки руководителей и команд
цифровой трансформации ВШГУ РАНХиГС



Малахов Александр Андреевич,
руководитель направления «Цифровое развитие» фонда
«Центр стратегических разработок»



Малыгин Владимир Вячеславович,
директор по развитию бизнеса Big Data в компании Tele2



Новичков Алексей Владимирович,
доцент факультета медиа, коммуникаций
и дизайна НИУ ВШЭ



Орлова Алиса Анатольевна,
редактор направления исследований и аналитики
Центра подготовки руководителей и команд цифровой
трансформации ВШГУ РАНХиГС



Петров Михаил Викторович,
директор департамента цифровой трансформации
Счетной палаты РФ, эксперт Центра подготовки
руководителей и команд цифровой трансформации ВШГУ
РАНХиГС, профессор департамента бизнес-информатики
Высшей школы бизнеса НИУ ВШЭ



Потапова Екатерина Геомаровна,
канд. филол. наук, руководитель направления
исследований и аналитики Центра подготовки
руководителей и команд цифровой трансформации
ВШГУ РАНХиГС



Саваровская Ольга Борисовна,
заместитель руководителя департамента
управления данными АНО «Аналитический центр
при Правительстве РФ»



Сазонов Виталий Анатольевич,
руководитель по инженерии данных ПАО «ВымпелКом»



Сергиенко Сергей Владимирович,
советник министра цифрового развития, связи и массовых
коммуникаций РФ, советник ФКУ «Государственные
технологии» (Гостех)



Скворцов Николай Владимирович,
канд. техн. наук, руководитель направления методологии
компании «Юнидата»



Сохоров Валерий Александрович,
начальник отдела управления данными департамента
цифровой трансформации аппарата Счетной палаты РФ



Стрекаловский Роман Степанович,
главный архитектор компании «Юнидата»



Суржко Денис Андреевич,
начальник управления перспективных алгоритмов машинного обучения департамента анализа данных и моделирования ПАО «Банк ВТБ»



Теплякова Дарья Олеговна,
редактор-переводчик направления исследований и аналитики Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС



Ткачева Ксения Андреевна,
директор Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС



Туманова Мария Витальевна,
аналитик Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС



Ушаков Дмитрий Вячеславович,
генеральный директор ООО «Гладиаторы ИБ»



Шамаев Юлий Анатольевич,
управляющий директор, управление перспективных алгоритмов машинного обучения департамента анализа данных и моделирования ПАО «Банк ВТБ»



Шклярук Мария Сергеевна,
академический директор Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС, генеральный директор АНО «Центр перспективных управленческих решений»

БЛАГОДАРНОСТИ

Центр подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС и авторский коллектив выражают признательность за помощь в работе над докладом:

Аношину Дмитрию Романовичу,
старшему инженеру данных Microsoft Gaming, Xbox

Белостоцкому Виктору Юрьевичу,
директору по технологиям (CDTO) Якутского республиканского онкологического диспансера

Берестневу Дмитрию Алексеевичу,
начальнику управления моделирования партнерств и ИТ-процессов департамента анализа данных и моделирования ПАО «Банк ВТБ»

Зориной Ольге Вячеславовне,
руководителю по развитию экосистемы ИНИД АНО «Центр перспективных управленческих решений»

Комину Михаилу Олеговичу,
директору по исследованиям АНО «Центр перспективных управленческих решений»

Никифорову Петру Владимировичу,
заведующему отделением рентгенохирургических методов диагностики Якутского республиканского онкологического диспансера

Ожаровскому Александру Вячеславовичу,
директору проектов и программ ПАО «Сбербанк», эксперту Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС

Полетаеву Олегу Викторовичу,
директору по развитию цифрового бизнеса группы «Интерфакс»

Цукарю Сергею Сергеевичу,
министру цифрового развития и связи Новосибирской области

ОНИ НАШЛИ УПРАВУ НА ДАННЫЕ

Вы готовы учиться управлять данными по-настоящему, с пользой для граждан и государства? Тогда предлагаем взглянуть на главные сюжеты работы с данными глазами опытных экспертов и авторов нашего навигатора, прежде чем двинуться в путь.



Глупо, что у кого-то в компании есть опись нашей офисной мебели, но ни у кого нет описи того, какими данными мы располагаем.

Даглас Лейни

Данные похожи на сырье. Оно полезно, но в необработанном виде непригодно для использования. Его необходимо преобразовать в газ, пластмассу, химикаты и тому подобные вещи...

Майкл Палмер

Data Science и продвинутая аналитика — это здорово, это модно, это круто и весело. Но это не волшебная таблетка и не серебряная пуля. Работа этих технологий зависит от степени достоверности информации. Мы не можем просто поверить в популярную технологию, про которую люди говорят: «Просто возьмите свои данные и положите их в озеро данных, и случится волшебство». Потому что это неправда, не так ли?

Кристофер Брэдли

Любой аспект данных, на который вы хотите обратить внимание с помощью диаграммы, может быть выражен посредством одного из пяти основных типов сравнения.

Джин Желязны

На уровне 1 наблюдается „феодализм данных“, на уровне 2 действует культура правил, но при этом существует „зоопарк информационных систем“. И только на уровне 5 возникает потребность в открытых данных, а организации начинают мыслить в долгосрочной перспективе.

Иван Добровольский



Руководить данными нужно для того, чтобы «делать правильные вещи» (doing the right things), а управлять данными нужно, чтобы «делать вещи правильно» (doing things right).

Джон Лэдли

Благодаря данным у госслужащих разного уровня появляется возможность принимать управленческие решения, руководствуясь не только опытом, который устаревает и вводит в заблуждение, не под влиянием чужого мнения, не интуитивно, а опираясь на актуальную и достоверную информацию.

Ксения Ткачева

Мы жестко увязываем результативность руководителя цифровой трансформации с решением острых массовых проблем. И в этом как раз заключается отличие от того, как аййтишники работали раньше.

Максут Шадаев

Работа с данными — в большой степени творческая. Пока у людей внутри не включится творчество, они реагируют на чужие идеи с подозрением: «пришли и хотят чего-то непонятного». Но потом появляется человек с аналитическим складом ума, который не просто выполняет задачи, а начинает думать, что можно изменить.

Михаил Петров

Интерес к доказательному подходу подтверждается тем, что в 2021 году Нобелевская премия по экономике присуждена экономистам Карду, Имбенсу и Ангристу за метод анализа причинно-следственных связей. Это одна из ключевых задач для госорганов, и мы надеемся, что когда-нибудь и российские ученые, работающие с российскими данными, тоже получат Нобелевку.

Мария Шклярчук

СОДЕРЖАНИЕ

ВВЕДЕНИЕ. КОМУ И ЧЕМ ПОЛЕЗЕН ЭТОТ НАВИГАТОР..... 12

БЛОК I ПОГРУЖАЕМСЯ В ТЕМУ

1. ИНФОРМАЦИОННЫЙ МУСОР ИЛИ НОВАЯ НЕФТЬ? 22

**1.1 Данные в госуправлении: инструмент,
а не предмет культа 23**

1.2 Данные как ресурс и как актив 25

**1.3 Теоретические основы управления данными
с позиции DAMA-DMBOK 28**

1.3.1 Назначение свода знаний 28

1.3.2 Одиннадцать областей знаний 29

1.3.3 Виды управления в области данных 32

2. КАКИЕ БЫВАЮТ ДАННЫЕ 34

2.1 Данные и информация в пирамиде знаний 35

2.2 Классификация данных по разным основаниям 40

2.3 Большие данные: три главных признака 46

3. КАК УСТРОЕНА РАБОТА С ДАННЫМИ В ГОССЕКТОРЕ..... 47

3.1 Особенности государственных данных 48

3.1.1 Какую роль играют данные в государстве 48

3.1.2 Национальная система управления данными 52

3.1.3 Откуда появляются и где хранятся госданные 55

3.2 Трудности в работе с государственными данными 58

3.2.1 Основные типы проблем 58

3.2.2 Причины проблем с данными и пути их решения 66

БЛОК II НАЧИНАЕМ РАБОТАТЬ

4. СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ 74

4.1 Жизненный цикл данных в DAMA-DMBOK 75

4.2 Создание системы управления данными в организации 77

4.3 Стратегическое планирование в области данных 80

4.4 Руководить, управлять, распоряжаться: в чем разница 81

4.5 Модель зрелости системы управления данными 83

5. ОРГАНИЗАЦИЯ УПРАВЛЕНИЯ ДАННЫМИ И РОЛИ В КОМАНДЕ..... 87

5.1 Кто отвечает за данные 88

5.2 Алгоритм создания эффективной команды 95

6. ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ ОБРАБОТКИ ДАННЫХ	96
6.1 Задачи инженера данных	97
6.2 Основные операции с данными: сбор, хранение, обработка	98
6.3 Подготовка к обеспечению процессов обработки данных в организации: чек-лист	102

БЛОК III ВЫСТРАИВАЕМ ПРОЦЕССЫ

7. ПРОБЛЕМА КАЧЕСТВА ДАННЫХ	104
7.1 Уровни зрелости качества данных в организации	105
7.2 Аудит управления данными. Бизнес-гlossарий и каталог	108
7.3 Управление качеством данных	111
7.4 Памятка для ответственного за качество данных	115
8. БЕЗОПАСНОСТЬ И РИСКИ ПРИ РАБОТЕ С ДАННЫМИ	116
8.1 Что угрожает данным	117
8.2 Свойства данных и виды рисков	119
8.3 Десять шагов к повышению безопасности	122

9. АНАЛИТИКА: ПРЕОБРАЗОВАНИЕ ДАННЫХ В ЗНАНИЯ И РЕШЕНИЯ	128
9.1 Базовые операции измерения	129
9.2 Уровни и результаты аналитики	130
9.3 Тренды анализа данных	133
9.4 Способы визуализации от графика до дашборда	135
9.5 Построение аналитики в организации	138
9.5.1 Выбор инструмента в зависимости от задачи	138
9.5.2 Подходы к построению структуры аналитики	142

ЗАКЛЮЧЕНИЕ. К ГОСУПРАВЛЕНИЮ НА ОСНОВЕ ДАННЫХ	145
---	------------

ПРИЛОЖЕНИЯ

А. ФРЕЙМВОРК «РЕШЕНИЕ ЗАДАЧ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ»	150
Б. ОПЫТ УПРАВЛЕНИЯ ДАННЫМИ В РОССИИ	157
В. ПРАВОВЫЕ ОСНОВЫ УПРАВЛЕНИЯ ГОСДАННЫМИ	178
ИСТОЧНИКИ ИЛЛЮСТРАЦИЙ	185
ГЛОССАРИЙ	186



ВВЕДЕНИЕ. КОМУ И ЧЕМ ПОЛЕЗЕН ЭТОТ НАВИГАТОР

 *Время чтения: 18 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Цифровая трансформация государства невыполнима без перехода на новый уровень принятия решений — на основе данных.
- › Чтобы принимать решения на основе данных, нужно освоить управление данными.
- › Умение работать с данными — часть квалификации госслужащего.
- › Данные нужны не сами по себе, а для достижения «человекоцентричных» целей.
- › Работа с данными дает возможность быстрее реагировать на изменения, создавать новые сервисы, снизить стоимость управления, принимать более обоснованные решения во всех отраслях.
- › Ряд госорганизаций уже активно выстраивает системную работу с данными. В ближайшей перспективе в госорганах появятся руководители по работе с данными (CDO).

Глобальные изменения, называемые цифровой трансформацией (ЦТ), имеют прямое отношение к управлению данными, отметил в выступлении на просветительском марафоне «Новое знание» председатель Правительства РФ Михаил Мишустин: «Этапы цифровой трансформации включают организацию сбора, хранения, передачи данных, обмен данными между людьми и устройствами, принятие решений»¹.

Цифровая трансформация государства невыполнима без перехода на новый уровень принятия решений — на основе данных, а он, в свою очередь, невозможен без выстраивания процессов и улучшения навыков управления данными. Работа с данными дает государству возможность создавать новые сервисы, быстрее реагировать на события, снизить стоимость управления, принимать более обоснованные и прозрачные решения во всех отраслях.

Сейчас госорганы делают в этом направлении значительные шаги, которые раньше трудно было себе представить. Ряд госорганизаций меняет работу с данными, следит за их качеством, модернизирует процессы сбора данных и их анализа, создает каналы межведомственного обмена данными, использует дашборды и тепловые карты, проводит исследования на основе данных, участвует в дата-хакатонах. В ближайшей перспективе в госорганах появятся руководители по работе с данными (Chief Data Officers, CDO). При этом важно, чтобы государство не только становилось «мегахабом по агрегации данных»², но и совершенствовало свою работу в целом, используя данные для повышения качества своих решений.



Для создания моделей будущих сервисов и проверки гипотез Роструд (см. приложение Б3) инициировал хакатон с привлечением молодых исследовательских команд. На основе их разработок были составлены предложения по модернизации сервиса.

Осваивать эту область знаний непросто и госсектору, и бизнесу. Так, 85% компаний заявляют, что хотят улучшить использование данных для принятия решений, но при этом 91% считает эту задачу сложной³. В каждой организации есть данные, которые могут быть полезны, но не всегда понятно, с чего начать работу с данными и как именно с ними работать. Важная часть преобразований — это изменение отношения к данным, повышение культуры работы с ними, что требует серьезных

Авторы:



К. Л. Киселева



А. А. Малахов



М. В. Петров



В. А. Сазонов



К. А. Ткачева

¹ Мишустин заявил, что цифровая трансформация ведет к экономике совместного потребления // ТАСС. URL: <https://tass.ru/ekonomika/12281237>

² Интервью TAdviser: Глава Минцифры Максуд Шадаев — о новых подходах к цифровизации государства // TAdviser. URL: https://www.tadviser.ru/index.php/Статья:Интервью_TAdviser:_Глава_Минцифры_Максуд_Шадаев_-_о_новых_подходах_к_цифровизации_государства

³ Sharma S. Data Is Essential To Digital Transformation // Forbes. URL: <https://www.forbes.com/sites/forbestechcouncil/2020/12/03/data-is-essential-to-digital-transformation/>

усилий от руководителя организации. Для построения всего здания data-driven госуправления требуются как собственно строители — управленцы с соответствующей квалификацией, так и фундамент — технологии и навыки управления данными, соответствующая инфраструктура и знания. «Все будет бессмысленно без людей, — отмечает академический директор Центра Мария Шклярук. — Если госслужащие не будут уметь работать с данными, понимать, что такое качество данных и строить решения на их основе, мы не сможем продвинуться с организационными и техническими моментами».

Что же касается требуемого фундамента, то очень надеемся, что этот навигатор станет одним из «кирпичей» в нем. В навигаторе показано, в том числе на актуальных примерах, как использовать работу с данными для изменения жизни людей.



Если бы у этого навигатора был слоган, он звучал бы так: «Данные — не ради данных, а ради человека». Данные полезны не сами по себе, а для достижения «человекоцентричных» целей, для качественной аналитики, надежных прогнозов и решения повседневных проблем граждан. Задача государства — выстроить работу с данными исходя из потребностей граждан, использовать данные на благо всех членов общества.

Почему руководителям государственных органов, государственным и муниципальным служащим нужно ориентироваться в теме данных? Назовем четыре главные причины.

1. Умение работать с данными — это часть квалификации госслужащего

Работу с данными можно считать неотъемлемой частью многих профессий, и госслужащие не исключение. Работодатель (а государство сейчас становится довольно требовательным работодателем) всегда выбирает тех, у кого есть актуальные компетенции.



В требования к кандидатам на должность заместителя руководителя федерального органа исполнительной власти (ФОИВ), ответственного за ЦТ⁴, включены:

- › знание и трансляция культуры и этики принятия решений на основе данных;
- › знание и понимание возможностей инструментов принятия решений на основе данных и искусственного интеллекта для цифровой трансформации бизнес-процессов и взаимодействия с гражданами;
- › опыт обеспечения безопасности данных и их соответствия нормативным требованиям.

⁴ О направлении требований к кандидатам на должность заместителя руководителя федерального органа исполнительной власти, ответственного за цифровую трансформацию, и методических рекомендаций по проведению тестирования их управленческих навыков. URL: <https://www.tadviser.ru/images/4/4f/05022020-2file.pdf>

Кроме того, для органов власти сейчас предусмотрен ряд ключевых показателей, для достижения которых требуется именно работа с данными. Среди множества нормативно-правовых актов (НПА) и документов разного уровня можно выделить приказ Минцифры⁵ № 600 и постановление Правительства⁶ № 542, описывающие целевые показатели ЦТ госуправления и ориентированные в том числе на сбор и анализ данных.

Еще один пример правительственного документа, где содержится требование квалифицированной работы с данными, — поручение заместителя председателя Правительства РФ Дмитрия Чернышенко от 29 июля 2020 года, согласно которому необходимо включить в ведомственные программы ЦТ (ВПЦТ) описания решений, основанных на применении искусственного интеллекта (ИИ), а также проектов, направленных на формирование массивов данных (датасетов⁷), пригодных для обучения на них ИИ-решений.

Помимо этого, от госслужащих все чаще требуется подтверждать результаты своей работы достижением определенных показателей, участвовать в рейтингах. Если в организации корректно выстроена работа с данными, ей будет проще собирать и предъявлять данные в нужный момент. Все вместе это дает портрет совершенно нового типа государственного управленца — высококлассного специалиста, профессионально владеющего навыками работы с данными.



В процессе внедрения системы управления данными Счетной палате РФ пришлось не только переформатировать процесс сбора данных, создать новую инфраструктуру и ввести новые роли, например институт дата-стюардов, но и много заниматься изменением культуры работы с данными, помочь сотрудникам осознать значимость управления данными (см. приложение Б2). Для сотрудников, работающих с данными, были созданы профили компетенций, проведено обучение, они приглашаются к участию в конференциях и хакатонах.

2. Без данных невозможна цифровая трансформация

Данные являются одним из основных ресурсов, позволяющих осуществить ЦТ; по фундаментальному влиянию на ЦТ они превосходят техническую инфраструктуру. Место данных в модели цифровой трансформации⁸ показано на рис. 1.

⁵ Приказ Минцифры России от 18.11.2020 № 600 (ред. от 14.01.2021) «Об утверждении методик расчета целевых показателей национальной цели развития Российской Федерации „Цифровая трансформация“» // КонсультантПлюс. URL: http://www.consultant.ru/document/cons_doc_LAW_372437/

⁶ Постановление Правительства Российской Федерации от 03.04.2021 № 542 «Об утверждении методик расчета показателей для оценки эффективности деятельности высших должностных лиц (руководителей высших исполнительных органов государственной власти) субъектов Российской Федерации и деятельности органов исполнительной власти субъектов Российской Федерации, а также о признании утратившими силу отдельных положений постановления Правительства Российской Федерации от 17 июля 2019 г. № 915» // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001202104130046>

⁷ Здесь и далее мы используем термины «датасет», «набор данных» и «массив данных» как синонимы.

⁸ Модель создана по материалам Сбера, см.: Государство как платформа: люди и технологии / под ред. М. С. Шклярук. М.: РАНХиГС, 2019. URL: <https://www.ranepa.ru/images/News/2019-01/16-01-2019-GovPlatform.pdf>

Культура и взаимодействие
Культура, ориентированная на взаимодействие и принципы честного и открытого общения, способствует развитию ЦТ

Люди и компетенции
Уникальные специалисты, способные быстро обучаться новым навыкам, чтобы успешно развиваться в эпоху ЦТ

Инфраструктура и инструменты
Мощные, надежные и гибкие системы и инструменты создают инфраструктуру, эффективно содействующую цифровой трансформации



Оптимизация процессов
Мониторинг и постоянная оптимизация процессов в соответствии с передовыми трендами ЦТ

Данные
Стратегический подход к управлению данными с обеспечением всестороннего доступа в режиме реального времени, обеспечение безопасности данных

Модели
Широкомасштабное применение моделей, основанное на непрерывном процессе инноваций

Рисунок 1. Модель цифровой трансформации

Традиционно данные использовались для подготовки отчетов и дескриптивной аналитики. Однако сейчас растет спрос на более продвинутые формы работы с данными, например на их применение для оперативного принятия решений в режиме реального времени, для построения прогностической и предписывающей аналитики (см. раздел 9.2), которая помогает руководителю ответить на вопросы «Что произойдет?» и «Что нужно сделать?».



Внедрение системы бизнес-аналитики в МФЦ Новосибирской области (см. приложение Б7) позволило в реальном времени автоматически определять отклонения в работе операторов. В результате повысилось качество услуг и сократилось время ожидания. Источниками данных служат системы оперативного учета (АИС МФЦ, электронная очередь и контакт-центр).

От лиц, принимающих решения (ЛПР), помимо мониторинга общей ситуации требуется принятие решений на основе данных. Именно так построена работа Координационного центра правительства⁹. Данные существуют не сами по себе, а в привязке к конкретным объектам и явлениям; профессионально организованная работа с данными позволяет трансформировать предприятие, организацию, отдельную отрасль и госуправление в целом.

Следующим важным шагом для госорганизации становится в таком случае применение принципов **доказательной политики** — государственной политики, основанной на строго установленных объективных доказательствах. Важным аспектом доказательной политики является

⁹ Образован постановлением Правительства РФ 12 февраля 2021 года. URL: <http://publication.pravo.gov.ru/Document/View/0001202102220029> Подробнее об этом органе см.: Координационный центр правительства Российской Федерации. URL: <http://government.ru/department/611/about/>

использование строгих научных исследований, которые невозможно реализовать без релевантных качественных данных, соответствующих процессов управления данными и навыков работы с данными¹⁰.

3. Данные помогают решать повседневные задачи

Широкое внедрение измерительной и контрольной аппаратуры (датчиков, сенсоров), применение методов машинной обработки, развитие государственных информационных систем привели к накоплению больших объемов данных в сфере госуправления, и пора начать добывать из этого «сырья» что-то полезное. Актуальной стала задача эффективного применения разнообразных данных. Примеры успешного решения этой задачи есть и в мире, и в России.

Если у руководителя и его команды есть соответствующий опыт и инструменты, многие проблемы, с которыми граждане сталкиваются ежедневно и которые годами пытаются решить органы власти, решаются быстрее и эффективнее.



В Новгородской области разработан и запущен цифровой сервис, предотвращающий возникновение очагов аварийности на дорогах¹¹. Специалисты регионального министерства цифрового развития проанализировали данные о дорожных элементах, повышающих безопасность («лежачих полицейских», пешеходных ограждениях, камерах видеофиксации), о количестве ДТП, числе погибших и раненых, о расстоянии от зданий до остановок общественного транспорта. Для получения данных были, в частности, переведены в машиночитаемый вид контракты на выполнение дорожных работ. На этой основе была создана система предиктивной аналитики, позволяющая прогнозировать количество ДТП в зависимости от наличия того или иного дорожного элемента, интенсивности транспортного потока и т. п. Реализация проекта позволит в ближайшие 2–3 года снизить аварийность и сэкономить бюджетные средства, которые идут на устранение последствий ДТП.

Кроме того, управление данными внутри организации упрощает сложные операции, повышает эффективность труда сотрудников, освобождает их от механической работы, помогает избежать переработок и авралов.

4. Данные позволяют принимать обоснованные решения

Чем выше уровень руководителя, тем сложнее ему управлять системой. Растет количество людей в подчинении, процессов, показателей, проектов, бюджетов. Данные и инструменты аналитики этих данных делают систему более управляемой.

¹⁰ Портал, посвященный доказательной политике в государстве, см.: ЦПУР. URL: <https://ebpm.cpur.ru/>

¹¹ В проекте применялся граф дорожной сети OpenStreetMap; компания «Яндекс» предоставила сведения о скорости транспортных средств из сервиса «Яндекс.Пробки». Для обработки данных разработчик использовал сервис MapBox. См.: Новгородская область: данные против ДТП // Вестник цифровой трансформации СIO.RU. URL: <https://cio.ru/articles/190221-Novgorodskaya-oblast-dannye-protiv-DTP>

Почему руководителю важно разбираться в качестве данных и методах их получения и обработки? Чтобы ему не показывали «удобную» статистику, «удобные» данные, по которым все хорошо, чтобы он не оказался в иллюзорном мире, доверяя некачественным или неполным данным. Данные позволяют тому, кто принимает решения, не полагаться, например, на чужие мнения или отрывочные сведения из разных источников, а видеть картину целиком, поддерживать связь с постоянно меняющейся ситуацией в организации и вокруг нее.

На уровне государства прозрачность и обоснованность государственных решений, в том числе решений о расходовании бюджетных средств, повышают доверие граждан, и это касается не только масштабных проектов, но и решений на местном уровне. Каждому гражданину важна уверенность в том, что выбранное решение, касающееся государства в целом, его города или улицы, — обоснованное, лучшее из возможного в конкретной ситуации. Поэтому доверие к данным и к аргументам, основанным на данных, меняет отношение граждан к действиям государства в целом.



Данные мобильных операторов могут использоваться для решения самых разных задач муниципального управления, например для оптимизации туристического потока или для планирования городской транспортной сети (см. приложение Б5). Использование данных позволяет не только установить точное число участников мероприятия, но и составить их портрет. Тем самым можно планировать мероприятия с учетом потребностей участников.

Перечисленные выше четыре причины делают очевидной потребность значительной части сотрудников государственных организаций в пособии, которое позволило бы им лучше разобраться в теме данных. Оставалось выбрать основные сюжеты и структурировать информацию так, чтобы она была доступна и полезна разным группам внутри нашей аудитории.

В начале работы над навигатором мы провели онлайн-анкетирование более 200 государственных и муниципальных служащих (руководителей и специалистов), сотрудников подведомственных учреждений и смежных организаций. Примечательно, что 80% участников выборки так или иначе уже работают с данными (непосредственно или через смежные проекты, см. рис. 2).

При этом разные категории госслужащих и сотрудников учреждений нуждаются в разных видах информации. Руководители высшего звена более, чем выборка в целом, заинтересованы в практической информации — в советах относительно организации работы с данными и принятия решений на основе данных. Руководители проектов хотели бы учиться на опыте других государственных органов и учреждений, у них сильнее выражен интерес к российским кейсам. Линейные руководители

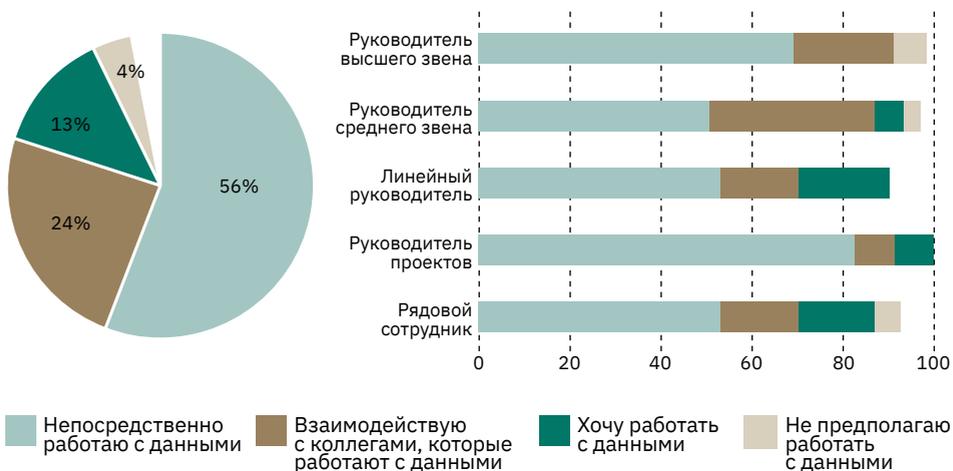


Рисунок 2. Степень текущей вовлеченности в работу с данными в зависимости от должности

интересуются современными методами анализа данных и способами обработки данных для машинного обучения и систем ИИ.

Мы постарались учесть эти запросы при выборе тем и кейсов для навигатора. Управление на основе данных — это конечная цель любой работы с данными. Но чтобы всерьез обсуждать технологии, практики, проблемы и вызовы, связанные с управлением на основе данных, нужно сначала освоить основополагающие, в некотором смысле азбучные принципы. Поэтому предлагаемый нами навигатор служит пособием именно в области управления данными: это тот фундамент, на котором строится принятие решений на основе данных и без которого системная работа в этой области невозможна. Он написан максимально понятным языком и содержит как вводные теоретические положения, связанные с управлением данными, так и «живую» практику работы с данными в государственных и муниципальных органах.

Что вы найдете в навигаторе?

Навигатор состоит из трех основных блоков и приложения.

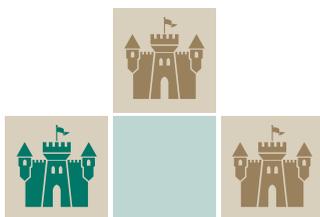
БЛОК 1. Погружаемся в тему (время чтения: 2 часа). Дает общее представление о данных и описывает базовую терминологию, классификации данных, проблемы в области российских государственных данных и возможные подходы к их решению. Этот блок будет полезен всем без исключения читателям доклада. Возможно, его будет достаточно тем читателям, кто хочет составить самое первое и общее представление о теме и у кого нет времени на большее (например, руководителям, сотрудникам тех команд, которые напрямую не работают с данными).

БЛОК 2. Начинаем путь (время чтения: 1 час). Описывает первые шаги, необходимые для начала работы с данными в организации: основные

подходы к созданию целостной системы управления данными, разработку дата-стратегии и архитектуры данных, принципы подбора команды и выстраивания инфраструктуры. Адресат этого блока – в первую очередь руководитель цифровой трансформации (РЦТ), который отвечает за старт этой работы в организации, а также CDO и в некоторых случаях члены его команды.

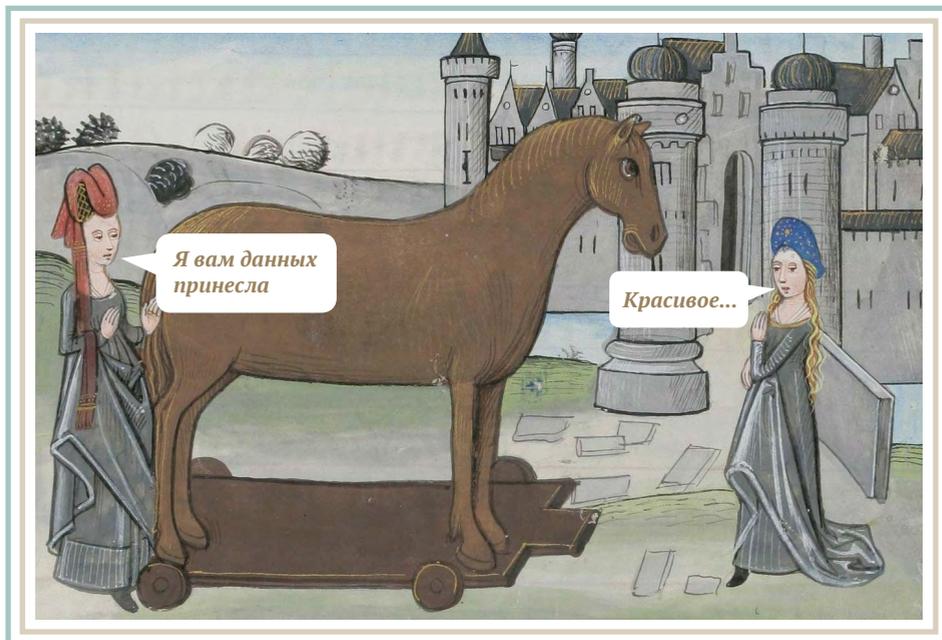
БЛОК 3. Выстраиваем процессы (время чтения: 1 час 30 мин.). Посвящен тем процессам, без которых невозможна работа с данными: обеспечению качества данных, их безопасности, созданию системы аналитики в организации. Этот блок особенно полезен начинающим CDO и членам их команд, но также и руководителям ЦТ и всем, кто интересуется темами качества данных и их анализа.

ПРИЛОЖЕНИЯ включают разработанный нами учебный фреймворк «Решение задач с использованием данных», позволяющий организации пройти самодиагностику для понимания своих сильных и слабых сторон в работе с данными, кейсы успешной работы с данными в российском госсекторе, а также обзор правовых основ управления государственными данными.



БЛОК I

ПОГРУЖАЕМСЯ В ТЕМУ



1. ИНФОРМАЦИОННЫЙ МУСОР ИЛИ НОВАЯ НЕФТЬ?

 *Время чтения: 25 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Данные, в зависимости от того, как устроена работа с ними, могут быть как основой прорывных решений, так и информационным мусором — если они разрозненны, некорректно собраны и не управляются должным образом.
- › Мало управлять данными как ресурсом — важно еще управлять ими как активом, то есть рассматривать как источник экономической выгоды.
- › Большинство организаций еще не ввели практику последовательного управления информационными активами, как это сделано в отношении активов материальных или финансовых; в частности, нет согласованности между управлением ИТ и управлением данными.
- › Наиболее известный свод знаний в области управления данными — DMВOK (Data Management Body of Knowledge). Его задача — дать полное и актуальное введение в дисциплину управления данными.
- › В DMВOK выделено 11 областей знаний по управлению данными — от качества данных до их архитектуры.
- › Управление данными — это разработка, выполнение и контроль выполнения политик, программ и практик предоставления, проверки, защиты и повышения ценности данных на протяжении всего их жизненного цикла.

1.1 ДАННЫЕ В ГОСУПРАВЛЕНИИ: ИНСТРУМЕНТ, А НЕ ПРЕДМЕТ КУЛЬТА

Принятие решений на основе данных — основное свойство государственного сектора, ориентированного на данные (Data-Driven Public Sector). Согласно программному документу Организации экономического сотрудничества и развития (ОЭСР) «На пути к формированию государственного сектора, ориентированного на данные»¹², в таком государстве (см. рис. 3):

- ▶ данные признаются ключевым стратегическим активом, определяется их ценность и измеряется их влияние;
- ▶ принимаются активные меры по уничтожению препятствий, мешающих управлять данными, обмениваться ими и использовать их повторно;
- ▶ данные используют, чтобы разрабатывать и отслеживать применение нормативно-правовой базы и предоставление государственных услуг;
- ▶ высоко ценится публикация открытых данных и обмен данными между государственными организациями и их подразделениями;
- ▶ соблюдаются права граждан, касающиеся данных (данные используются этично, прозрачно, обеспечено право на приватность и безопасность данных).

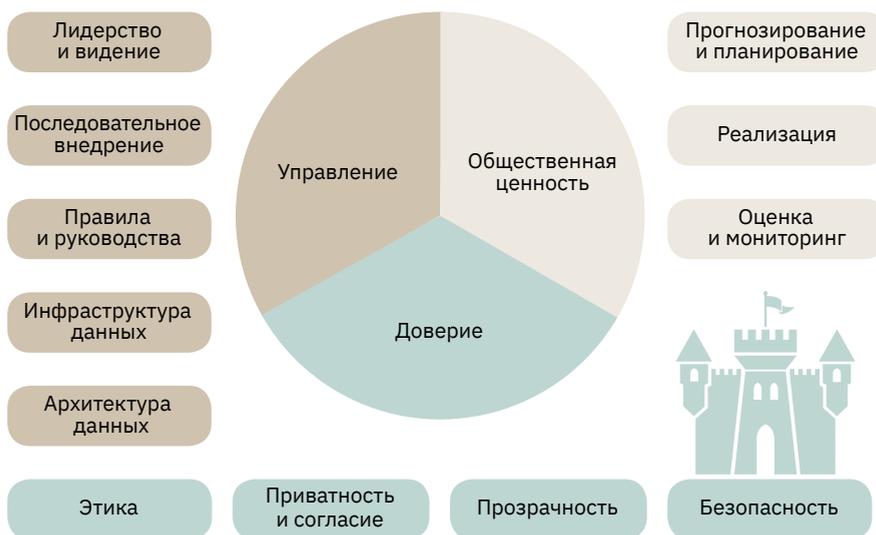


Рисунок 3. Государство, ориентированное на данные

¹²The Path to Becoming a Data-Driven Public Sector // OECD iLibrary. URL: <https://www.oecd-ilibrary.org/sites/059814a7-en/index.html?itemId=/content/publication/059814a7-en>

Авторы главы:



И. Д. Димитров



Е. Г. Поталова



Н. В. Скворцов



Д. О. Теплякова

Страны, внедрившие стратегический подход к использованию данных в госсекторе, лучше прогнозируют потребности общества и, следовательно, разрабатывают более эффективные долгосрочные планы. Активное использование данных играет важную роль при анализе всех видов государственной деятельности, позволяя их совершенствовать. В госуправлении на основе данных выше уровень прозрачности и подотчетности, а это способствует доверию к власти, вовлечению общества в процесс изменений и повышению результативности госуправления во всех секторах. Доказательная политика, основанная на данных (Evidence-Based Policy), требует проверки реальной эффективности любых мер, предложенных правительством, будь то меры по снижению бедности, по повышению качества образования, природоохранные меры.



«Все эти вопросы стали обсуждаться не только и не столько на основе политических или идеологических аргументов, сколько на основе того, можно ли объективно проверить, что мера, предложенная правительством, <...> достигла того конечного эффекта, которое правительство заявляло, когда начинало ее реализовывать и выделяло на нее финансирование»¹³.

**Мария Шклярчук, академический директор
Центра подготовки РКЦТ**

Но одно лишь внедрение в каждое министерство ответственных за данные, сложных инструментов и аналитики нужного результата не даст. От чего еще зависит результат?

- › Критически важно **качество данных**, непрерывное обеспечение которого невозможно без использования специальных технологий и привлечения специалистов.
- › Нужно квалифицированно и разумно ставить цели сбора и **анализа данных**. Например, если на основе данных предполагается изучить все варианты развития событий в той или иной области, только прогноза будет недостаточно: должна быть модель развития ситуации.
- › Данные должны быть актуальными и **постоянно обновляться** (при этом старые, «исторические» данные не должны уничтожаться или модифицироваться задним числом), потому что прогнозирование и принятие решений происходит в динамике.

Недостаточно сделать базу, купить систему, внедрить технологию, чтобы работа с данными дала результат. Без учета перечисленных и многих других факторов управление данными происходит формально, внедренные инструменты не используются, а сами данные превращаются в предмет

¹³ Из выступления М. С. Шклярчук «ЦПУР как драйвер доказательной политики в России». URL: <https://ebpm.cpur.ru>

карго-культ¹⁴. Карго-культ возникает там, где участники процессов действуют не вдаваясь в суть дела, только ради отчетности. Например, приложение для удаленного общения с врачом — полезная технология, но в ней нет смысла в той местности, где у 80% жителей нет связи.

При построении информационной системы (ИС) государственная организация исходит из того, что первоочередная задача — бороться с хаосом в данных. Однако автоматизация деятельности нередко приводит к парадоксальному увеличению хаоса: например, в процессе автоматизации поставили программный комплекс, который не использует предыдущие данные, но производит новые, так что порядка становится меньше. Важно уметь правильно поставить задачу аналитикам и исследователям данных, а затем проконтролировать ее выполнение. В фокусе внимания должны быть не наличие и объем данных, а их обработка и качество¹⁵, — и главное, задачи, которые могут быть решены на основе этих качественных данных.



В начале работы с данными все организации сталкиваются с вопросами показателей и измерений. Возникает опасность ради хороших показателей утратить не только логику работы с данными, но и весь смысл деятельности организации. Известны, например, случаи, когда для того, чтобы не портить показатели, медики отказывали в госпитализации пациентам, которых сложно оперировать, или не отправляли на вызов машину скорой помощи.

1.2 ДАННЫЕ КАК РЕСУРС И КАК АКТИВ

В 2006 году бизнесмен Клайв Хамби ввел в обращение метафору «Данные — это новая нефть», а вице-президент американской Ассоциации рекламодателей (Association of National Advertisers) Майкл Палмер развил эту 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...»¹⁶ Одновременно другие эксперты писали о том, что аналогия между данными и нефтью не работает, что данные могут быть токсичными или даже «мусорными»¹⁷.

Любая организация должна стремиться ко всестороннему и глубокому пониманию своих ресурсов и их стратегическому анализу, однако кейсов успешного управления данными как ресурсом пока не так много.

¹⁴ Термин, обозначающий поверхностную имитацию чего-либо (например, процесса) без глубинного понимания причинно-следственных связей в надежде получить результаты, аналогичные тем, которые дает объект копирования. Карго-культ является проявлением магического мышления. Подробнее см.: Скоренко Т. Культ карго: «Кока-кола», посланная богами // Мир фантастики. URL: <https://www.mirf.ru/science/kult-kargo/>

¹⁵ См. анализ тенденции к росту значимости качества данных по частоте упоминания в резюме, росту инвестиций в эту отрасль и другим параметрам: Data Quality Unpacked // Gradient flow. URL: <https://gradientflow.com/data-quality-unpacked/>

¹⁶ Palmer M. Data is the New Oil // ANA Marketing Maestros. URL: https://ana.blogs.com/maestros/2006/11/data_is_the_new.html

¹⁷ См., например: James J. Data as the new oil: The danger behind the mantra // The Enterprisers Project. URL: <https://enterpriseproject.com/article/2019/7/data-science-data-can-be-toxic>; Карелов С. Данные — это новый мусор // Эксперт. URL: <https://expert.ru/2020/08/6/dannyie---eto-novyyij-musor/>



Один из субъектов РФ решал задачу улучшения качества данных в едином государственном реестре недвижимости (ЕГРН), касающихся кадастровой стоимости объектов — это самая важная статья для начисления налогов на недвижимость. Для этого нужно было найти ошибки и исправить данные. Поиск ошибок осуществлялся в том числе путем анализа экстремумов (например, у однокомнатной квартиры указана площадь 3000 кв. м — вероятно, в числе была потеряна запятая). Далее подразделение Росреестра занималось исправлением этих ошибок. Сложность состояла в том, что данные в ЕГРН являются подтверждением юридически значимого факта, а значит, просто внести в них правки невозможно. Но все-таки значительное количество найденных технических ошибок (как в примере с площадью квартиры) удалось исправить в особом порядке на основе закона о ЕГРН. В итоге после улучшения качества баз данных начисление налогов в субъекте РФ стало более справедливым и увеличилось на несколько миллиардов рублей¹⁸.

Ресурсы по отдельности не создают конкурентного преимущества; они должны быть задействованы все вместе, только тогда они формируют так называемую **организационную способность**¹⁹ — способность организации осуществлять специфическую деятельность²⁰. Оргспособности в модели Гранта (см. рис. 4) требуют, чтобы опыт сотрудников организации был

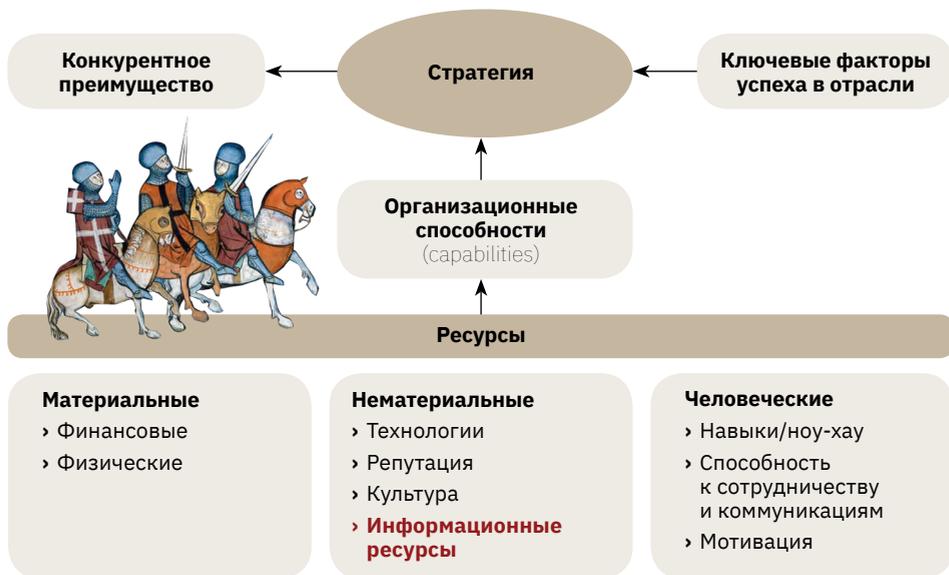


Рисунок 4. Взаимосвязь между ресурсами, способностями и конкурентным преимуществом

¹⁸ Кейс предоставлен О. В. Саваровой.

¹⁹ Термин capabilities переводится в русских изданиях как «способности» и «возможности». Мы вслед за одним из ведущих российских экспертов в области стратегического управления В. С. Катъкало предпочитаем первый вариант. См. также: Катъкало В. С. Место и роль ресурсной концепции в развитии теории стратегического управления // Вестник С.-Петербургского ун-та. Сер. Менеджмент. 3. С. 3–17. URL: <https://cyberleninka.ru/article/n/mesto-i-rol-resursnoy-kontseptsii-v-razvitiy-teorii-strategicheskogo-upravleniya-predisloviye-k-razdelu-1>

²⁰ Грант Р. М. Современный стратегический анализ. 5-е изд. / пер. с англ.; под ред. В. Н. Фунтова. СПб.: Питер, 2008.

объединен с оборудованием, технологией и другими ресурсами. Для получения конкурентного преимущества на основе оргспособности она должна быть уникальна и/или должна обеспечивать ключевые факторы успеха (например, помогать создавать ценности для пользователей).



Данные, будучи ценным ресурсом, могут одновременно стать источником специфических проблем. В отличие от других ресурсов, они могут быть легко скопированы, переданы множеству людей, а затем использованы десятками способов (см. раздел 8.2). Организациям трудно обеспечить соблюдение конфиденциальности и других политик, касающихся безопасного использования данных, что создает репутационные и финансовые риски.

ИТ-подразделения (если перед ними не поставлены специальные задачи по работе с данными) смотрят на данные в первую очередь с точки зрения затрат: данными необходимо управлять, чтобы минимизировать затраты на поддержание этого ресурса. Но мало управлять данными как ресурсом — **важно еще управлять ими как активом**²¹, то есть рассматривать как источник экономической выгоды. Разница во взглядах на данные как ресурс и как актив показана на рис. 5.



Рисунок 5. Данные как ресурс и как актив

²¹ Под активом (asset) понимается имеющийся в собственности организации или контролируемый ею экономический ресурс, содержащий в себе или производящий ценность. При этом он может быть монетизирован или из него может быть извлечена иная выгода. Понятие экономической выгоды, хоть и происходит из бизнеса, применимо к госсектору с точки зрения повышения производительности, удовлетворенности граждан полученными услугами и т. д.

В большинстве организаций еще не введена практика последовательного управления информационными активами, как это сделано в отношении активов материальных или финансовых. В частности, сохраняется несогласованность между управлением ИТ и управлением данными. Как отмечал авторитет в области управления данными Джон Лэдли²², «команды разработчиков поспешно производят сотни приложений и сервисов, не принимая во внимание вопросы использования связанных с ними данных».

Данные, в зависимости от того, как устроена работа с ними, могут быть как сырьем для прорывных решений в организации, так и мусором — если они неконсистентны, разрознены, некорректно собраны и не управляются должным образом. Поэтому в настоящем навигаторе сделан акцент на том, как наладить управление данными, чтобы противостоять негативным воздействиям и извлечь ценность из имеющихся информационных ресурсов, превратить их из «мусора» в «нефть» цифровой экономики.

Что нужно сделать организации, чтобы начать обращаться с данными как с активом? Некоторые идеи изложены в книге Дагласа Лейни²³ «Инфономика. Информация как актив: монетизация, оценка, управление».

1.3 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ УПРАВЛЕНИЯ ДАННЫМИ С ПОЗИЦИИ DAMA-DMBOK

1.3.1 НАЗНАЧЕНИЕ СВОДА ЗНАНИЙ

Ряд концептуальных понятий в области управления данными мы приводим на основании наиболее известного, полного и всеобъемлющего свода знаний в области управления данными²⁴ — DAMA-DMBOK. Этот документ создан Международной ассоциацией управления данными (Data Management Association International, DAMA), которая объединяет профессионалов в области управления данными.

Первое издание «Руководства DAMA к своду знаний по управлению данными» (Data Management Body of Knowledge, DMBOK) вышло в 2009 году. Эта книга заложила фундамент развития профессий, относящихся к управлению данными. В 2017 году опубликовано второе издание — DAMA-DMBOK2, а в 2020 году — его перевод на русский язык.

²² Джон Лэдли (John Ladley) — автор книг об управлении информацией и руководстве данными Making Enterprise Information Management (EIM), Work for Business: A Guide to Understanding Information as an Asset and How to Design, Deploy and Sustain an Effective Data Governance Program.

²³ Даглас Лейни (Douglas Laney) — специалист в области хранилищ данных и стратегического управления данными, автор модели зрелости процессов управления корпоративной информацией Gartner, а также один из авторов термина Big Data (большие данные).

²⁴ Сегодня существует только два полноценных комплекта справочных материалов по управлению данными, которые не зависят от поставщика соответствующих решений: DAMA-DMBOK и CMMI DMM (модель зрелости управления данными Института CMMI), см.: 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.

Задачи DAMA — собирать, систематизировать и пропагандировать лучший опыт, вырабатывать единую терминологию и фреймворки для знаний и подходов. Основное назначение DMBOK — предоставление полного, точного и актуального введения в дисциплину управления данными, в частности:

- ▶ выработка общепринятого согласованного представления об областях знаний по управлению данными²⁵;
- ▶ определение руководящих принципов управления данными;
- ▶ предоставление стандартных определений для наиболее часто используемых понятий (общих и по областям знаний);
- ▶ обзор лучших практик, методов и методик, а также альтернативных подходов;
- ▶ краткий обзор общих организационных и культурных вопросов;
- ▶ уточнение границ сферы управления данными;
- ▶ предоставление полезных общепотребительных фреймворков управления данными (см. разделы 4.2, 4.3).



Зачем DMBOK нужен «простому управленцу» или «простому госслужащему»? Подобные своды знаний есть и для других областей — от управления проектами (PMBOK²⁶) до бизнес-анализа и программной инженерии (BABOK²⁷ и SWEBOK²⁸ соответственно). Они полезны тем, что в концентрированной форме излагают основные идеи и принципы конкретной области, позволяя погрузиться в нее быстро и довольно глубоко. Конечно, максимум пользы из таких сводов знаний могут извлечь специалисты, имеющие опыт практической работы, но и начинающим они дают ясную картину предметной области. Мы не только рекомендуем ознакомиться с DAMA-DMBOK, но и даем в навигаторе ряд понятий в соответствии с этой методологией.

1.3.2 ОДИННАДЦАТЬ ОБЛАСТЕЙ ЗНАНИЙ

Управление данными согласно DAMA-DMBOK строится вокруг основного фреймворка²⁹ DAMA, включающего колесо DAMA, шестиугольник факторов среды и контекстные диаграммы. На рис. 6 представлено колесо DAMA — диаграмма областей знаний по управлению данными.

В DMBOK выделено 11 областей знаний по управлению данными.

1. Руководство данными (Data Governance) обеспечивает стратегическое управление информационными активами организации. Это одна из функций управления данными, главная по отношению ко всем другим

²⁵ В DMBOK области знаний по управлению данными иногда также называются функциями управления данными.

²⁶ Project Management Body of Knowledge («Свод знаний по управлению проектами»).

²⁷ Business Analysis Body of Knowledge («Свод знаний по бизнес-аналитике»).

²⁸ Software Engineering Body of Knowledge («Свод знаний по программной инженерии»).

²⁹ В русском переводе DAMA-DMBOK термину framework соответствует термин «рамочная структура»; в этой книге ради согласованности с другими главами (в частности, с главой 7) был выбран термин «фреймворк».

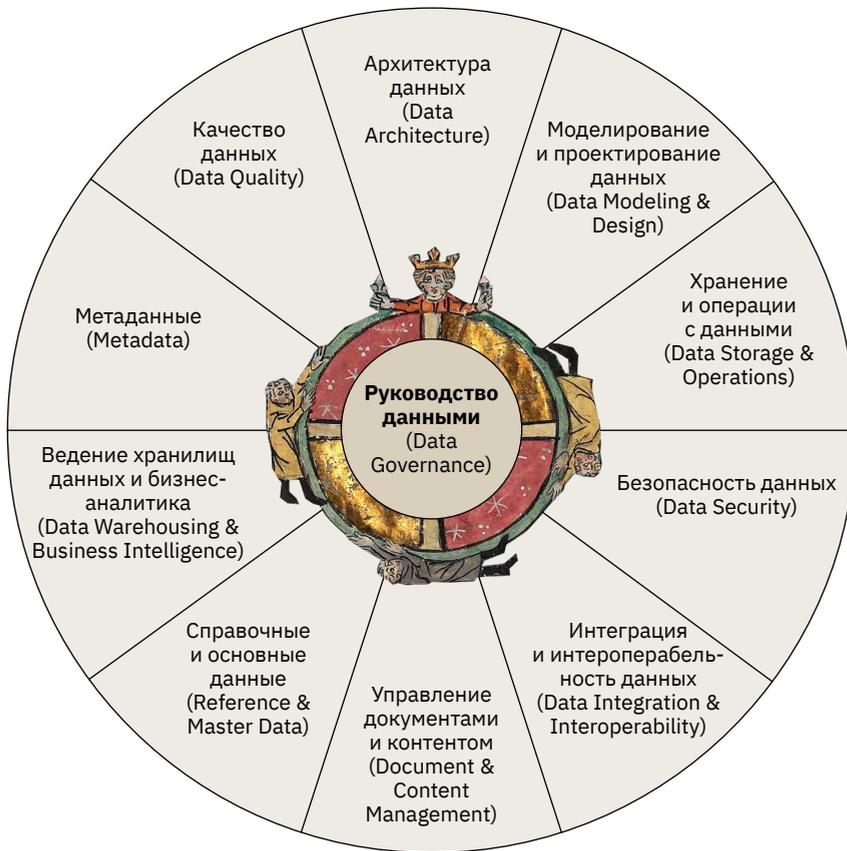


Рисунок 6. Колесо DAMA: области знаний по управлению данными

его функциям. Руководство данными необходимо, чтобы обеспечить такие процессы принятия решений по поводу данных, которые отвечают интересам организации.

2. Архитектура данных задает схему управления информационными активами в соответствии со стратегией организации, формулируя стратегические требования к данным и способы удовлетворения этих требований.

3. Моделирование и проектирование данных — это процесс выявления, анализа, фиксации и распространения требований к данным в точно определенной форме, называемой моделью данных. Процесс происходит итеративно и может включать разработку концептуальной, логической и физической моделей.

4. Хранение данных и операции с ними. Проектирование и внедрение решений для хранения данных с целью максимального увеличения их ценности. Операции обеспечивают сопровождение данных на протяжении всего их жизненного цикла от планирования до удаления.

5. Безопасность данных обеспечивает приватность и конфиденциальность данных, защиту от взлома, надлежащий доступ к ним.

6. Интеграция и интероперабельность данных включают процессы, связанные с перемещением и консолидацией данных как внутри хранилищ, приложений и организаций, так и между ними.

7. Управление документами и контентом включает планирование, внедрение и мониторинг процессов, которые связаны с управлением ЖЦ данных (подробнее см. главу 2), полученных из разнообразных неструктурированных источников, особенно документов, которые необходимы для обеспечения выполнения нормативных требований.

8. Справочные данные и основные данные (мастер-данные) включают непрерывную сверку и согласование основных совместно используемых данных с целью обеспечить последовательное использование во всех системах наиболее точных, актуальных и релевантных сведений о ключевых бизнес-единицах.

9. Ведение хранилищ данных и бизнес-аналитика предполагают планирование, внедрение и контроль процессов управления теми данными, на основе которых принимаются решения, и создание условий, в которых информационные работники (knowledge workers) смогут извлекать ценность из данных с помощью их анализа и формирования отчетности.

10. Метаданные предполагают планирование, проведение и контроль работ по обеспечению доступа к данным о данных, например, к определениям, моделям, потокам данных и другой информации, критически важной для понимания самих данных и систем, где они создаются, содержатся и где к ним обеспечивается доступ.

11. Качество данных предполагает разработку и внедрение методов управления качеством с целью измерять, оценивать и повышать пригодность данных к использованию внутри организации.

Руководству данными отведено центральное место в деятельности по управлению данными; именно оно обеспечивает согласованность и сбалансированность всех функций. Другие области (архитектура, моделирование и проектирование данных и т. д.) равномерно распределены вокруг центра колеса DAMA. Все области необходимы, чтобы управление данными можно было назвать зрелым, но реализовывать их можно постепенно и в том порядке, который отвечает нуждам организации. Проработка всех сегментов колеса позволяет построить полноценную систему управления данными — в этом состоит практический смысл этого инструмента.

Для каждой из областей знаний в DMBOK подробно описаны основные факторы, влияющие на эту область (см. рис. 7): это люди (их роли и ответственность), процессы и технологии.



Рисунок 7. Шестиугольник факторов среды

Наконец, контекстные диаграммы описывают для каждой области знаний цели, входные материалы (и их поставщиков), работы (и их участников), результаты (и их потребителей), а также методы, инструменты и метрики³⁰.



Колесо DAMA отражает набор областей знаний на верхнем уровне; шестиугольник факторов среды показывает общие структурные элементы этих областей на более низком уровне; контекстные диаграммы детально представляют устройство каждой области.

Изучение методологии DAMA-DMBOK и основных областей управления данными, составляющих колесо DAMA, может стать первым шагом на пути к полноценному управлению данными.

1.3.3 ВИДЫ УПРАВЛЕНИЯ В ОБЛАСТИ ДАННЫХ

Рассмотрим термины, применяемые к деятельности по управлению данными:

- › управление данными (Data Management, DM);
- › управление информацией (Information Management, IM);
- › управление корпоративной информацией (Enterprise Information Management, EIM);
- › управление информационными активами (Information Asset Management, IAM).

³⁰ См. подробную схему заполнения контекстной диаграммы для отдельной области знания в презентации выпускающего редактора DAMA DMBOK2 Сьюзан Эрли. URL: <http://dama-phoenix.org/wp-content/uploads/2015/09/DAMA-Phoenix-DMBOK2.pdf>

Согласно определению DMBOK³¹, **управление данными** — это разработка, выполнение и контроль выполнения политик, программ и практик предоставления, проверки, защиты и повышения ценности данных на протяжении всего их жизненного цикла. Поскольку в DMBOK термины «информация» и «данные» в применении к управлению данными с известными оговорками взаимозаменяемы (соотношение этих терминов будет подробно обсуждаться в главе 2), «управление данными» и «управление информацией» также описывают одну и ту же реальность. Джон Лэдли отмечает³², что рекомендации DMBOK для управления данными применимы как на локальном уровне (на уровне отдельных подразделений), так и на корпоративном (уровне организации в целом).

Управление корпоративной информацией подразумевает управление на уровне всей организации, причем под это определение подпадают не только коммерческие структуры, но и министерства, ведомства, любые большие организации.



Управление информацией и управление корпоративной информацией обычно рассматриваются как программы в смысле подхода Program Management³³, под программой понимается «совокупность взаимосвязанных проектов и другой деятельности, направленных на достижение общей цели и реализуемых в условиях общих ограничений»³⁴.

Важное отличие управления корпоративной информацией от управления данными вообще — подход к данным как к активу (см. раздел 1.2). В этой связи четвертый ключевой термин в области управления данными — управление информационными активами. Это набор принципов и концепций, в котором данные рассматриваются как стратегически важный актив организации³⁵. Управление информационными активами можно воспринимать как новый, датацентричный образ мышления³⁶.

Между управлением информационными активами и управлением корпоративной информацией существует неразрывная связь: первое поставляется концепции, а второе, будучи программой (см. выше), реализует эти концепции³⁷. Далее управление данными мы будем обсуждать именно в контексте управления корпоративной информацией.

³¹ В этой главе мы берем за основу определения из DAMA-DMBOK.

³² Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

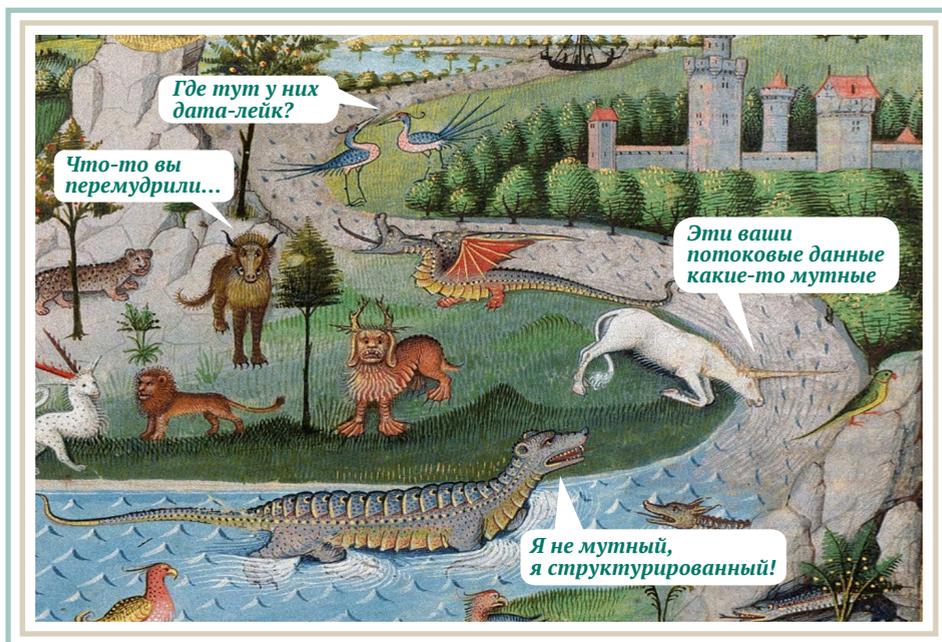
³³ Smith P., Edge J., Parry S., Wilkinson D. Crossing the Data Delta: Turn the data you have into the information you need // Entity Group Limited. 2–16. 2016.

³⁴ См. ГОСТ Р 54871-2011. Проектный менеджмент. Требования к управлению программой.

³⁵ В качестве примеров элементов IAM можно привести предложенные Джоном Лэдли «Общепринятые принципы учета информационных активов» (Generally Accepted Accounting Principles, GAAP), а также опубликованный в 2017 году «Лидерский манифест о данных». URL: <https://dataleaders.org/wp-content/uploads/2019/04/manifesto-russian.pdf>

³⁶ Основные аспекты IAM отражены в книге: Лейни Д. Инфономика. Информация как актив: монетизация, оценка, управление. М.: Точка, 2020.

³⁷ Там же.



2. КАКИЕ БЫВАЮТ ДАННЫЕ

 *Время чтения: 22 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Данные — это факты или наблюдения, неорганизованные и необработанные, не передающие никакого конкретного смысла и не имеющие ценности, потому что они лишены контекста и интерпретации.
- › Данные лежат в основе «пирамиды знаний»; в ней также есть информация (организованные, или структурированные, данные), знания и мудрость.
- › Основные проблемы возникают на этапе преобразования данных в информацию. Для преодоления разрыва между ними необходимо изменить отношение организации к ее данным: их следует воспринимать как стратегически важный актив и стремиться к повышению их ценности.
- › Данные могут классифицироваться по назначению и области применения, по структурированности, источникам и другим признакам.
- › Данным разных категорий соответствуют разные требования, присущи разные риски и отведены разные роли в организации. Например, основные данные имеют иное назначение и области применения, нежели транзакционные данные; соответственно, требования к управлению данными двух этих категорий будут отличаться.
- › К особой категории можно отнести большие данные (Big Data), которые принято определять через 3 V: объем, скорость, разнообразие.

2.1 ДАННЫЕ И ИНФОРМАЦИЯ В ПИРАМИДЕ ЗНАНИЙ

Применительно к управлению данными термины «данные» и «информация» не всегда последовательно различают, но в некоторых случаях разница существенна³⁸, что отчасти связано с их происхождением.



В английский язык слово *information* пришло в конце XIV века из французского (первоисточник — латинское *informatio* ‘разъяснение, истолкование, сообщение’), а с середины XV века за английским *information* закрепилось значение ‘переданные сведения, относящиеся к определенной теме’. Слово *data* в значении, близком к современному, стало использоваться в английском языке значительно позже. Оно происходит от латинского *datum* — ‘данная вещь’ (от глагола *dare* — ‘давать’).

В XVII веке благодаря быстрому распространению книгопечатания появилось множество научных книг, стремительно вырос объем совместно используемых сведений, и для обозначения таких сведений стали применять термин *data* (сначала он употреблялся в более узком значении ‘исходные факты для вычислений при решении математических задач’)³⁹.

Историческое смысловое соотношение между «данными» и «информацией» отчасти сохраняется и сегодня. Приведем перевод толкований из Оксфордского словаря современного английского языка⁴⁰:

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 другое подобное устройство.

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

При обсуждении соотношения понятий в этой области обычно используется аббревиатура DIKW (*Data — Information — Knowledge — Wisdom*, «данные — информация — знания — мудрость»). Это

Авторы главы:



А. С. Константинов



Н. В. Скворцов



Р. С. Стрекаловский

³⁸ См.: DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020; Лейни Д. Инфономика. Информатика как актив: монетизация, оценка, управление. М.: Точка, 2020.

³⁹ The History of Data // Circa Navigate. URL: <https://circa-navigate.corsairs.network/the-history-of-data-5663fc4f1398>

⁴⁰ Oxford Dictionary of Current English / Edited by Della Thompson. 2nd Edition. Oxford University Press, 1993. Следует заметить, что здесь в отношении термина «информация» мы ограничиваемся его нестрогими определениями и не обсуждаем научные определения (используемые в теории связи, теории информации, кибернетике и связанные с понятием количества информации); последние не имеют прямого отношения к вопросам управления информационными активами.

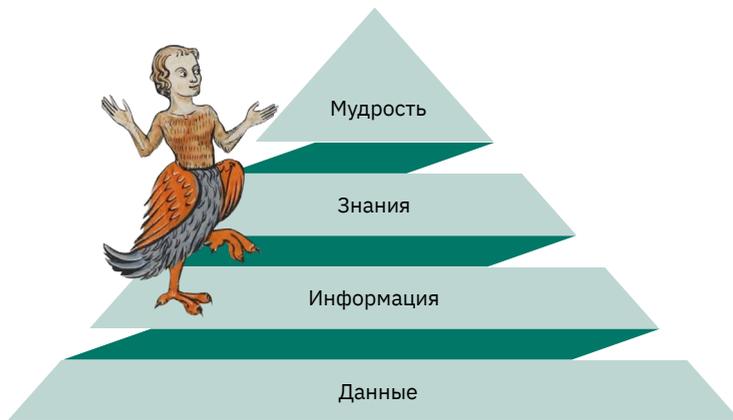


Рисунок 8. Пирамида знаний

соотношение часто изображают в виде пирамиды⁴¹ (рис. 8) и называют иерархией знаний, пирамидой знаний, информационной пирамидой и т. п.

Для более полного понимания каждого уровня пирамиды рассмотрим относящиеся к нему определения⁴². Определения данных обычно делают акцент на том, чего данным не хватает, а именно: они лишены смысла или ценности, не организованы и не обработаны.

Обобщенное определение может выглядеть так.

ДАННЫЕ — это дискретные, объективные факты или наблюдения, неорганизованные и необработанные, не передающие никакого конкретного смысла и не имеющие ценности, потому что они лишены контекста и интерпретации.

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

ИНФОРМАЦИЯ — форматированные данные, обработанные с определенной целью, которым придан смысл посредством добавления контекста.

Таким образом, **информация в большинстве случаев рассматривается как организованные или структурированные данные**. Обработка обеспечивает соответствие данных конкретной цели или контексту и тем самым делает их значимыми, ценными, релевантными.

⁴¹ Такое графическое представление впервые появилось в работе Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link // OceanTeacher: a training system for ocean data and information management. Intergovernmental Oceanographic Commission (UNESCO), 2004. URL: https://web.archive.org/web/20071202033948/http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DIKWchain.pdf

⁴² Для этого воспользуемся сведениями из обзоров известных монографий и руководств по информационным системам и управлению знаниями, в частности: Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy // Journal of Information Science. 2007. № 33 (2). P. 163–180. URL: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>; Van Meter Heather J. Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom // Law Technology and Humans. Vol. 2. No. 2 (November, 2020). P. 69–80. URL: <https://lthj.qut.edu.au/article/view/1470>. См. также упомянутую выше работу Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link.

До сих пор нет единого мнения о природе знаний: разные эксперты предлагают разные определения. В то время как данные отражают свойства вещей, знания относятся к способностям людей, предрасполагающим их действовать определенным образом. Мы предлагаем такое определение.

ЗНАНИЯ — это совокупность данных и информации (к которым добавляются экспертные мнения, опыт, другие знания), которая была организована и обработана с целью передачи понимания, накопленных результатов обучения и компетенции так, чтобы получился ценный актив, который можно применить в текущей деятельности для принятия решений.

Что касается определения мудрости, то его можно найти далеко не у всех авторов, описывающих иерархию DIKW. Мудрость — более расплывчатое понятие, чем знания; она сильнее связана с человеческой интуицией, пониманием, интерпретацией и действиями, со способностью трактовать и использовать знания, чем с информационными системами⁴³. Обобщая, можно предложить такое определение.

МУДРОСТЬ — это способность действовать наиболее подходящим образом с учетом того, что известно (знания) и что приносит наибольшую пользу (этические и социальные соображения).

В ближайшие годы область данных, вероятно, будет расти экспоненциально (см. раздел 2.3), но еще неизвестно, увеличатся ли в объеме знания и мудрость. Специалисты отмечают⁴⁴, что, хотя количество собираемых данных стремительно растет⁴⁵ (рис. 9), не все из них становятся «базой» для информации, не говоря уже о знаниях или мудрости.

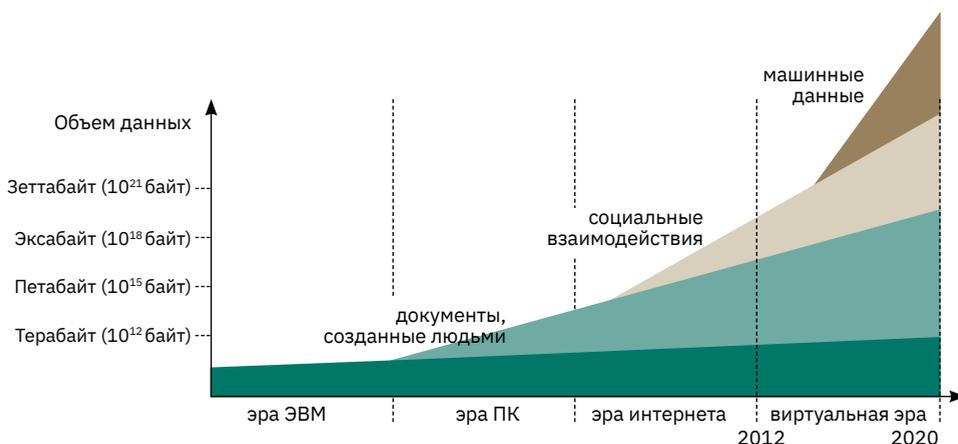


Рисунок 9. Экспоненциальный рост данных в последние десятилетия

⁴³ См. также: Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy // Journal of Information Science. 2007. № 33 (2). P. 163–180. URL: <http://www-public.imtbs-tsp.edu/~gibson/Teaching/Teaching-ReadingMaterial/Rowley06.pdf>

⁴⁴ См.: Van Meter Heather J. Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom // Law Technology and Humans. Vol. 2. No. 2 (November, 2020). P. 69–80. URL: <https://lthj.qut.edu.au/article/view/1470>

⁴⁵ Data Management & Analytics: Legacy Challenges and Technological Trends // DataArt. URL: <https://www.dataart.com/blog/data-management-analytics-legacy-challenges-and-technological-trends>

Данные могут быть неточными или ложными; пока неизвестно, остается ли их удельный вес постоянным по мере роста общего объема собираемых данных или доля некачественных данных увеличивается. Это может происходить, например, из-за стремительного роста разного рода фейковых новостей или из-за отрицательной реакции людей на сам сбор данных. Если наша цель — обретение знаний и мудрости (а не построение и обслуживание дата-центров), то по-настоящему ценным будет увеличение количества имеющихся у нас точных и достоверных данных.



Для чего профессионалу, работающему с данными, нужно получить представление о пирамиде знаний? Она не только имеет теоретическую и методологическую ценность, но и дает понимание того, что данные — это основа, из которой должны вырастать более зрелые уровни управления государством, те уровни, где присутствуют знания и мудрость, проявляясь в обоснованном принятии решений, использовании доказательной политики, в переходе к более ответственному и прозрачному управлению.

Интересно, что уровни пирамиды знаний могут быть сопоставлены с классами информационных систем, которые преимущественно используются в работе с ними (рис. 10).

Важное свойство пирамиды знаний — постоянные преобразования одного уровня в другой: данных — в информацию, информации — в знания и т. д. Это свойство имеет прямое отношение к деятельности компаний, организаций и государства в целом. Деятельность любой организации



Рисунок 10. Сопоставление уровней пирамиды знаний с классами ИС



Рисунок 11. Цикл преобразований «данные — информация — знания»

можно представить как множество циклических цепочек «данные — информация — знания»⁴⁶ (рис. 11). В этом цикле данные 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 результате операции по продаже клиенту товара или по предоставлению гражданину государственной услуги появляются данные, связанные с этой операцией. Далее они могут быть преобразованы в ценную информацию (в нашем примере это будет, в частности, информация о том, что для определенного клиента заказ выполнен повторно). Эта информация поступает в хранилище, содержащее корпоративные знания, что позволяет на основании совокупности сведений о заказах сделать определенные выводы (допустим, вывод о том, что на недавно проведенную рекламную кампанию лучше реагируют клиенты старше 40 лет). Исходя из этих знаний организация может предпринять действия — в нашем случае провести кампанию, направленную на клиентов в возрасте до 40 лет, и тем самым увеличить продажи. Соответственно, появятся новые данные, новая информация и новые знания, на основе которых будут предприняты новые действия, и цикл повторится.

⁴⁶ См. об этом: 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. Redman T. C. Data Driven: Profiting from Your Most Important Business Asset. Harvard Business Review Press, 2008. Мы не включаем в эту цепочку мудрость в связи с ее уже упомянутым неопределенным статусом.

Практика большинства организаций показывает, что **преобразование данных в информацию является слабым звеном в цепочке**. Очень часто мы наблюдаем разрыв между данными и информацией. Причиной могут быть данные: плохо определенные, разрозненные, дублирующиеся, низкого качества или устаревшие. Их может быть недостаточно или же, наоборот, слишком много. Для преодоления разрыва организации следует изменить отношение к своим данным, а именно рассматривать их как стратегически важный актив (см. раздел 1.2), повышать их ценность и устранять слабые места.

Несмотря на упомянутый разрыв между данными и информацией, они тесно переплетены и по отдельности не существуют. Данные — это форма информации, а информация — это форма данных⁴⁷.

При этом внутри организации полезно проводить четкую границу между информацией и данными хотя бы для донесения требований и ожиданий по различным направлениям практической работы до сотрудников⁴⁸.



«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]».

Предметом управления являются и данные, и информация; при этом качество того и другого возрастает лишь при согласованном управлении ими с учетом потребностей конечных пользователей.

2.2 КЛАССИФИКАЦИЯ ДАННЫХ ПО РАЗНЫМ ОСНОВАНИЯМ

Организация, которая не знает, какими данными располагает и какие могла бы получить, не может использовать их в качестве актива.



Даглас Лейни приводит высказывание директора по информационным технологиям крупной страховой компании: «Глупо, что у кого-то в компании есть описание нашей офисной мебели, но ни у кого нет описи того, какими данными мы располагаем»⁴⁹.

⁴⁷ Это, в частности, отмечается в: DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

⁴⁸ С этой и другими базовыми практиками цифровой трансформации, с основами проектного управления и сквозными технологиями помогут познакомиться открытые курсы Центра подготовки руководителей и команд цифровой трансформации. См.: «Эпоха цифрового развития: основы цифровой трансформации» (URL: <https://stepik.org/course/65359/>) и «Цифровая трансформация. Быстрый старт», (URL: <https://stepik.org/course/89635/>)

⁴⁹ Лейни Д. Инфономика. Информация как актив: монетизация, оценка, управление. М.: Точка, 2020.

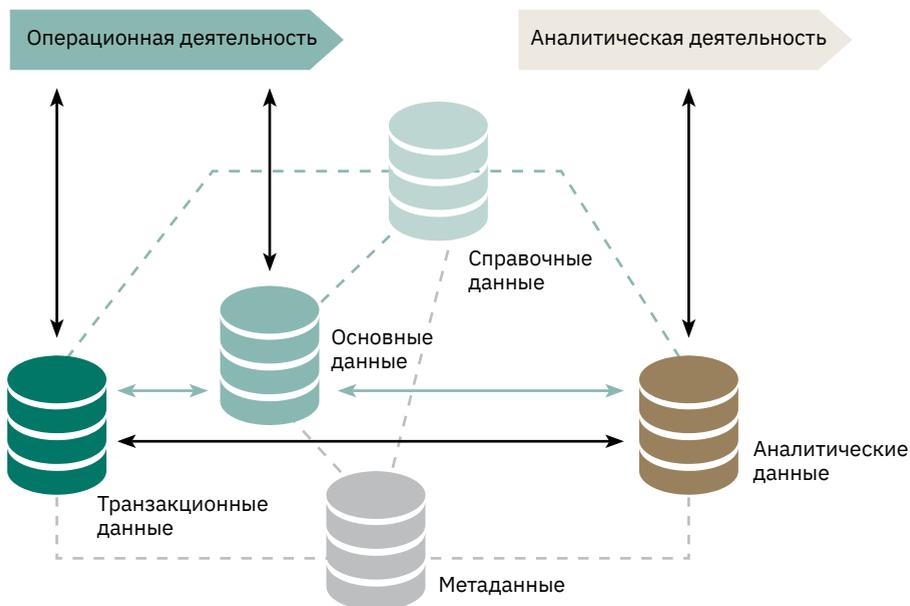


Рисунок 12. Взаимосвязи основных категорий данных в деятельности организации

Первый и самый очевидный способ начать разбираться с данными — это выделить и определить разные типы данных. Специалисты по управлению данными обычно классифицируют их по нескольким основаниям в зависимости от того, какие задачи им нужно решить. Так, по **назначению и области применения** принято выделять:

- › **метаданные**, описывающие структуру и характеристики данных;
- › **справочные данные** — данные из справочников, международных, общероссийских и отраслевых классификаторов и т. п.;
- › **основные данные** — данные об объектах и бизнес-сущностях, представляющих ценность для организации (о клиентах, продуктах, работниках, технологиях и материалах и т. п.)⁵⁰;
- › **транзакционные данные** — сведения, отражающие результат изменения данных, относящиеся к фиксированному моменту времени, не изменяющиеся в будущем.

Часто в отдельную категорию выделяют аналитические данные (см. рис. 12): они образуются из основных, справочных и транзакционных данных и используются в аналитической деятельности организации.

На рис. 13 в качестве примера показаны роли каждой из категорий данных в информационном обеспечении процессов организации (в данном случае коммерческой). Отметим фундаментальную роль справочных и основных

⁵⁰ Очень часто вместо термина «основные данные» применяется термин «мастер-данные». Преимущества термина «основные данные» в том, что он определен и используется в комплексе национальных стандартов ГОСТ Р ИСО 8000 «Качество данных». См., в частности, ГОСТ Р ИСО 8000-2-2019 «Качество данных. Часть 2. Словарь». URL: <https://docs.cntd.ru/document/1200169126>

Основные данные (Запись о клиенте)

№ клиента	5001	Тип	03 Коммерческий
Наименование клиента	ООО «Радуга»		
Адрес	Российская Федерация, Алтайский край		
Индекс	659635		

Справочные данные (Тип клиента)

Код	Тип
01	Государственный
02	Муниципальный
03	Коммерческий
04	Сфера образования

Метаданные

Имя поля	Определение	Тип поля	Длина поля
Номер товара	Уникальный номер товара, присвоенный менеджером	Числовое	15
Наименование товара	Краткое наименование товара, указываемое в едином товарном перечне	Символьное	30

Транзакционная запись (Заказ на доставку)

№ заказа	1234	Тип клиента	03 Коммерческий
№ клиента	5001	Адрес доставки	Российская Федерация, Алтайский край
Наименование клиента	ООО «Радуга»	Индекс	659635
Номер товара	50223	Наименование товара	Сумка для ноутбука
		Количество	4
		Цена	1000 р.
		Итоговая цена	4000 р.

Основные данные (Запись о товаре)

Номер товара	50223
Наименование товара	Сумка для ноутбука

Справочные данные (Цена)

Номер товара	Тип клиента	Цена
50223	01	950
50223	02	950
50223	03	1000
50223	04	900

Рисунок 13. Роли отдельных категорий данных в информационном обеспечении процессов организации

данных (мастер-данных) и важность их качества. Например, при наличии ошибок в данных о номере товара или типе клиента цена доставки будет определена некорректно (см. связи, обозначенные пунктиром), что может привести к серьезным финансовым последствиям.

По степени структурированности можно выделить:

- › **структурированные данные** — данные, имеющие строго фиксированную структуру, определяемую формальной моделью данных (например, реляционной схемой);
- › **полуструктурированные (слабоструктурированные) данные** — данные, не имеющие строго определенной структуры, но предполагающие наличие правил, позволяющих выделять отдельные семантические элементы при их интерпретации, прежде всего правил расстановки тегов и других маркеров, отмечающих и выделяющих элементы данных (например, файлы, созданные с использованием языка XML и его многочисленных производных, html-страницы и др.);
- › **неструктурированные данные** — данные, произвольные по форме, не имеющие строго определенной структуры и не организованные по определенным правилам.

На рис. 14 приведены примеры форматов хранения и передачи данных для каждой из категорий.

Отдельно следует сказать о машинных и потоковых данных. К **машинным данным** относится информация, автоматически генерируемая компьютерами, датчиками и иными устройствами без вмешательства человека (когда мы говорим об интернете вещей или о данных медицинского мониторинга, имеются в виду как раз машинные данные). Машинные данные становятся одним из основных источников информации, это в первую очередь относится к данным контроля и аудита (то есть к сведениям, фиксируемым в различных журналах регистрации).

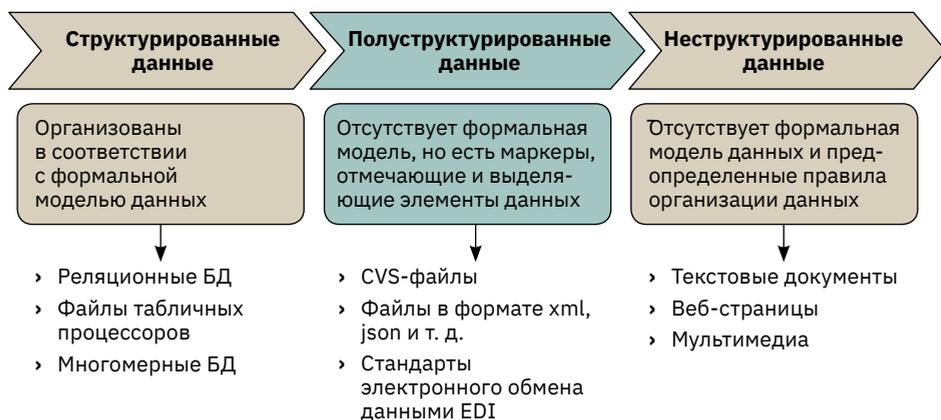


Рисунок 14. Форматы хранения и передачи данных с разной степенью структурированности

Потоковые данные могут относиться почти к любой из перечисленных выше категорий, однако у них имеется одно дополнительное свойство. Данные поступают в систему при возникновении некоторых событий, а не загружаются в хранилище данных большими массивами. Примером потоковой обработки данных является сервис YouTube, проводящий анализ данных пользователей исходя не только из просмотренных ими полностью видеозаписей и трансляций, но и из материалов, которые они пропустили. Другим примером могут служить данные телеметрии, полученные с любого датчика или набора датчиков (например, системы «умный дом»).

При инвентаризации информационных активов целесообразно выделить группы в зависимости от **источников поступления** данных. Приведем в качестве примера классификацию информационных ресурсов коммерческой организации, предложенную экспертами компании Gartner⁵¹.

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. Чаще всего их собирают с помощью датчиков и мониторинга процессов предприятий. Их источником могут быть, например, кассовые аппараты, подключенные к банковской системе, интеллектуальные счетчики, голосовая связь.

«Темные данные» организации не собирают и не хранят специально; они возникают (попутно) при ведении бизнеса и взаимодействии с сетевыми сервисами, а затем остаются в интернет-архивах и локальных хранилищах. К ним относятся электронные письма, мультимедиа, системные журналы.

Публичные данные распространяются госорганами и коммерческими компаниями. Их ценность раскрывается в совокупности с другими источниками; они позволяют определить направления социально-экономического развития в отдельном городе, в стране или в группе стран.



Категория открытых данных тесно связана с категорией публичных данных. Термин «открытые данные» возник в американском научном сообществе в 1995 году как призыв к свободному обмену данными. Принципиальное отличие публичных данных состоит в том, что их использование регулируется законом — доступ к ним возможен, например, по специальному запросу. Смысл открытых данных в другом: они должны быть опубликованы еще до того, как кому-то понадобятся. Подробнее об открытости данных см. раздел 3.1.3.

Коммерческие данные — это сведения, представляющие коммерческий интерес, которые открыто размещаются в цифровой среде, в частности данные об активах, размещенные на открытых площадках.

⁵¹ Лейни Д. Инфономика. Информация как актив: монетизация, оценка, управление. М.: Точка, 2020.



Рисунок 15. Соотношения между категориями данных

Данные крупных социальных сетей (Facebook, «ВКонтакте», Twitter, Instagram и др.) активно используются как государственными, как и коммерческими структурами для получения ценной информации о рыночных и социальных тенденциях, о продуктах, услугах и сервисах, об общественном мнении и т. д.

Вернемся к общей картине управления данными. На рис. 15 показаны в общем виде отношения между категориями данных в организации⁵².

Данные из категорий, расположенных сверху, как правило, являются базовыми; они участвуют в формировании данных нижних категорий. **Поэтому по мере движения вверх по списку категорий требования к качеству соответствующих данных возрастают.** Также по мере продвижения вверх по списку категорий увеличивается продолжительность жизненного цикла данных. При движении вниз по списку категорий увеличивается объем самих данных, а также частота их изменений.

Поскольку данным разных категорий соответствуют разные требования, присущи разные риски и отведены разные роли в организации, многие инструменты управления данными сфокусированы на проблемах конкретных категорий данных. Например, основные данные имеют иное назначение и области применения, нежели транзакционные данные; соответственно, требования к управлению данными двух этих категорий будут отличаться.

⁵² См.: 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.

На диаграмме объединены категории, выделяемые по разным признакам (назначению и степени структурированности); метаданные, справочные, основные и транзакционные данные являются преимущественно структурированными.

2.3 БОЛЬШИЕ ДАННЫЕ: ТРИ ГЛАВНЫХ ПРИЗНАКА

К особой категории можно отнести большие данные (Big Data). Этот термин связан с техническими аспектами сбора и обработки данных; он не предполагает конкретные виды данных. Смысл термина Big Data значительно шире, чем просто указание на большой объем данных: он указывает также на высокую скорость поступления данных и разнообразие источников и форматов получаемой информации.

Big Data принято определять по трем «V»⁵³.

- › **Volume** (объем) данных представляет собой количество данных, доступных для анализа с целью извлечения полезной информации. Ключевым для развития технологий обработки больших данных стал рост объемов данных вследствие интернет-активности. Например, на YouTube загружается 300 часов видео каждую минуту, а объем мобильного трафика достиг 6,2 млрд гигабайт в месяц.
- › **Velocity** (скорость) обработки данных — это скорость потока создания, хранения, анализа и визуализации данных. Быстрота изменений приводит к необходимости обработки большого количества данных за короткий промежуток времени. Так, Google обрабатывает 40 тысяч поисковых запросов в секунду — то есть 3,5 млрд запросов в день.
- › **Variety** (разнообразие) данных заставляет анализировать данные разных типов из различных предметных областей. Кроме того, большие данные включают в себя не только структурированные, но полуструктурированные и неструктурированные данные — и последних большинство. К большим данным относятся аудио- и видеофайлы, изображения, данные постов в социальных сетях и другие текстовые форматы, данные о переходе по ссылкам, машинные данные, данные датчиков.

ГОСТ⁵⁴ выделяет еще одну, четвертую ключевую характеристику:

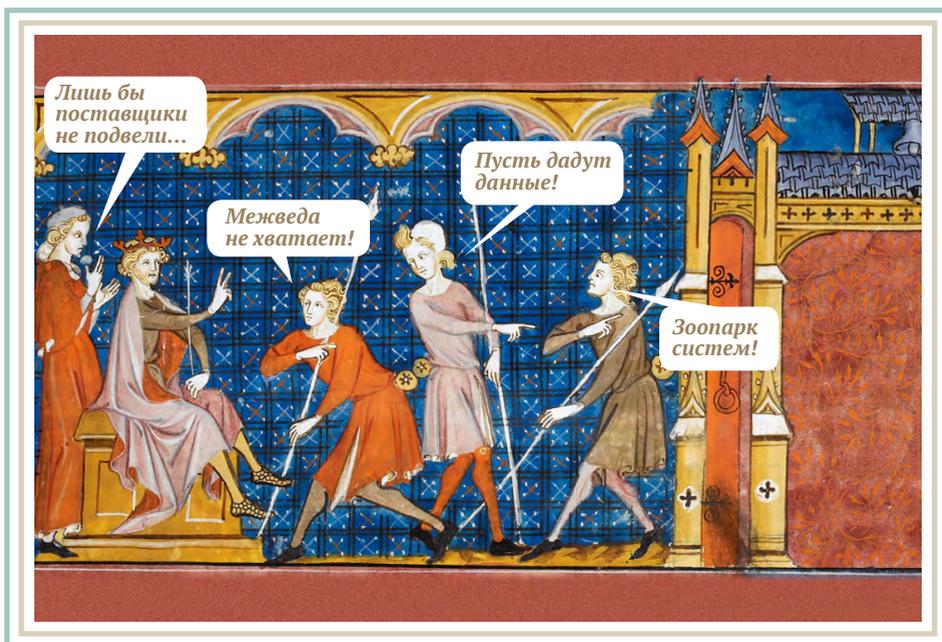
- › **Variability** (вариативность) данных — это изменения в скорости передачи данных, их формате и (или) структуре, семантике и (или) качестве, которые влияют на работу с данными. Вариативность приводит к необходимости реорганизации архитектур, интерфейсов, методов обработки, влияет на интеграцию, слияние, хранение, применимость и использование данных.

Иногда к первым четырем V добавляются еще шесть дополнительных признаков: veracity (достоверность), visualization (визуализация), validity (валидность), vulnerability (уязвимость), volatility (волатильность) и value (ценность)⁵⁵.

⁵³ Лейни Д. Инфономика. Информация как актив: монетизация, оценка, управление. М.: Точка, 2020.

⁵⁴ ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии. Большие данные. Обзор и словарь». Введен в действие с 1 ноября 2021 года.

⁵⁵ The 10 Vs of Big Data // TDWI. URL: <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx> Дополнительные признаки описывают уже не характеристики больших данных, а скорее проблемы, возникающие в связи с их использованием.



3. КАК УСТРОЕНА РАБОТА С ДАННЫМИ В ГОССЕКТОРЕ

КЛЮЧЕВЫЕ ИДЕИ

- ▶ К госданным относятся данные, которыми распоряжаются и за которые отвечают органы власти и организации госсектора: создают, собирают, хранят, защищают, изменяют, удаляют.
- ▶ Отдельно стоит выделить юридически значимые сведения в реестрах. Они нужны для оказания госуслуг, в том числе проактивных, и адресной помощи.
- ▶ К качеству госданных предъявляются жесткие требования, поскольку это фундамент управления на основе данных. Их роль в принятии решений на всех уровнях власти будет только расти.
- ▶ Разобщенность государственных данных выражается в ряде проблем с качеством, получением доступа, форматами на внутриведомственном и межведомственном уровнях, а также между федеральными властями, субъектами федерации и муниципалитетами.
- ▶ В части межведомственного обмена данными активную роль будет играть Национальная система управления данными (НСУД), которая должна собрать сведения обо всех госданных и обеспечить их единообразие, транзакционную и аналитическую доступность.
- ▶ Чем чаще данные используются в разных сферах, тем больше на них обращают внимание, тем они лучше, качественнее и удобнее.



О. А. Виниченко



А. А. Малахов



О. Б. Саваровская

3.1 ОСОБЕННОСТИ ГОСУДАРСТВЕННЫХ ДАННЫХ

 *Время чтения: 27 мин.*

3.1.1 КАКУЮ РОЛЬ ИГРАЮТ ДАННЫЕ В ГОСУДАРСТВЕ

Государство опирается в своей работе на данные, и в этом отношении оно ничем принципиально не отличается от любой крупной организации. Для государства характерны все операции с данными, составляющие их жизненный цикл: создание, сбор, обработка, публикация, хранение, удаление, обеспечение защиты и т. п. Во всех сферах, где работают с данными, к ним применимы одни и те же принципы, инструменты, методы оценки их значимости.

К **ГОСУДАРСТВЕННЫМ ДАННЫМ**⁵⁶ относятся все данные, которыми распоряжаются органы власти и организации государственного сектора, — которые они создают, собирают, хранят, защищают, изменяют, удаляют и за которые отвечают на всем жизненном цикле работы с данными⁵⁷.

Государство использует данные при анализе, подготовке, принятии управленческих решений (см. также раздел 1.1). Для поддержки принятия решений нужна как совокупность первичных данных, так и результаты их анализа — разного рода отчетность, статистические и аналитические выкладки, причем при отсутствии одного из слагаемых качество управленческих решений резко снижается.

Отдельно стоит выделить такие важные госданные, как юридически значимые сведения в реестрах. Они способны подтверждать юридические факты о людях и их документах, о движимом и недвижимом имуществе, организациях, начиная с самых привычных, таких как фамилия, имя, отчество и дата рождения конкретного человека. Данные в реестрах очень нужны для оказания разнообразных государственных услуг и мер поддержки, а в будущем потребуются для проактивного оказания услуг⁵⁸ и адресной помощи.

⁵⁶ Говоря о государственных данных, мы имеем в виду обращение данных внутри государства, так как вопросы трансграничной передачи данных и их использования, например, международными организациями или другими странами требуют отдельного рассмотрения.

⁵⁷ Концепция создания и функционирования национальной системы управления данными (утверждена распоряжением Правительства Российской Федерации от 03.06.2019 № 1189-р, с изм. на 14.05.2021) дает следующее определение: «Государственные данные — информация, содержащаяся в информационных ресурсах органов и организаций государственного сектора, а также в информационных ресурсах, созданных в целях реализации полномочий органов и организаций государственного сектора». URL: <https://docs.cntd.ru/document/554802572>

⁵⁸ Проактивное (упреждающее) оказание госуслуг означает, что ведомство или цифровой госсервис предлагают человеку доступную ему услугу, не дожидаясь его обращения. При этом ведомство само запрашивает и получает необходимые документы и справки благодаря налаженному межведомственному обмену. Гражданину остается лишь согласиться на предложенную услугу (хотя он может и отказаться), и она будет оказана.



Единый государственный реестр недвижимости (ЕГРН) Росреестра — первый в России электронный информационный ресурс, электронные записи в котором (в реестрах ресурса) имеют юридическую значимость.

Потребителей госданных можно разделить на большие группы⁵⁹. Каждая включает несколько подгрупп, использует госданные с разными целями для решения разных задач (см. табл. 1). Депутатам законодательных собраний данные нужны для обоснования законопроектов, руководителям — для подкрепления цифрами документов стратегического планирования, правительству РФ и органам власти субъектов РФ — для принятия решений, затрагивающих миллионы граждан. Автоматизация государственного надзора и контроля, использование ИИ при принятии и верификации решений — перспективы госуправления в ближайшем будущем. Поскольку в результате деятельности государства возникают значимые для общества решения, к госданным предъявляются жесткие требования. Ошибки в данных бизнеса опасны для бизнеса и для его клиентов, но их масштаб, как правило, меньше, чем в госданных. Кроме того, от ошибок в госданных страдает значительно больше людей и государство в целом.

Госданные необходимы государству еще и для того, чтобы оценивать качество госуправления, достижение стратегических и тактических целей, эффективность деятельности госслужащих, в первую очередь руководителей министерств и высших должностных лиц регионов. Их работа может быть признана неэффективной при недостижении поставленных показателей, а выполнение задач оценивается на основе анализа все большего количества данных⁶⁰.



В приказ Минцифры № 600 включены методики расчета пяти целевых показателей, в том числе «достижение „цифровой зрелости“ ключевых отраслей экономики и социальной сферы» и «увеличение доли массовых социально значимых услуг, доступных в электронном виде, до 95 процентов»⁶¹.

Источниками информации для расчета уровня цифровой зрелости названы административные данные заинтересованных ФОИВ и государственных внебюджетных фондов РФ, формируемые в рамках автоматизированных ИС.

⁵⁹ Группы соответствуют разным целевым аудиториям внутри государства. Мы рассматриваем группы исходя из распределения целевых аудиторий внутри государства, как было сказано ранее. Список групп разработан на основании описаний, созданных в Минэкономразвития при участии АНО «ЦПУР». См.: Минэкономразвития провело стратегическую сессию по качеству российских открытых данных // Минэкономразвития. URL: https://www.economy.gov.ru/material/news/minekonomrazvitiya_provelo_strategicheskuyu_sessiyu_po_kachestvu_rossiyskih_otkrytyh_dannyh.html

⁶⁰ Постановление Правительства РФ от 03.04.2021 № 542 «Об утверждении методик расчета показателей для оценки эффективности деятельности высших должностных лиц <...> субъектов Российской Федерации и деятельности органов исполнительной власти субъектов Российской Федерации <...>» предполагает работу с данными для достижения заданных показателей.

⁶¹ Приказ Минцифры России от 18.11.2020 № 600 (ред. от 14.01.2021) «Об утверждении методик расчета целевых показателей национальной цели развития Российской Федерации „Цифровая трансформация“» // КонсультантПлюс. URL: http://www.consultant.ru/document/cons_doc_LAW_372437/

Таблица 1. Цели использования госданных разными группами потребителей

Потребители госданных	Представители группы	Цели использования госданных
Государство	Федеральные и региональные органы исполнительной власти, муниципалитеты, органы законодательной власти, суды, правительство, президент, государственные институты развития, территориальные органы управления и др.	<ul style="list-style-type: none"> › Принятие управленческих, законодательных и судебных решений на основе анализа данных в соответствии с принципами доказательной политики и для совершенствования регуляторных решений › Формирование подробных и полных представлений о сферах применения государственных программ и политик › Повышение качества госуправления (в т. ч. изменение законодательства, оказание госуслуг, выделение госфинансирования, снижение нагрузки на объекты аудита, развитие технологий) › Контроль расходов и оптимизация бюджетов, снижение издержек › Управление рисками в части контрольно-надзорной деятельности и проектной деятельности
Бизнес	Технологические компании, ИТ-стартапы, малый бизнес, ритейл, производство и др.	<ul style="list-style-type: none"> › Управление рисками (такими как регуляторные решения, колебания валютных курсов, изменение ставок госпошлин для импортеров и другими) › Анализ рынка и конкурентов › Поиск новых возможностей (освоение новых рыночных ниш, разработка новых продуктов, в том числе ИТ-решений, повышение качества существующих услуг и продуктов) › Оптимизация расходов, получение льгот при производстве, снижение издержек
Наука и образование	НИИ, вузы, аналитические агентства, частные лица, занимающиеся наукой и исследованиями, и др.	<ul style="list-style-type: none"> › Проведение исследований, поиск закономерностей и тенденций, выявление неочевидных причинно-следственных связей › Определение новых, перспективных, малоизученных сфер исследований › Изучение общества и глобальных процессов в нем на больших временных отрезках для предсказания изменений › Публикация научных статей и развитие научного знания › Создание и разработка новых технологий
Общество, СМИ, НКО	Общественные организации, СМИ и НКО как представители гражданского общества, специализирующиеся на обеспечении гражданского контроля деятельности государства, граждане (физические лица) и др.	<ul style="list-style-type: none"> › Получение достоверной информации и обеспечение общественного контроля › Анализ ситуации, организация и проведение общественно-политических кампаний › Реализация своих непосредственных функций, в частности выдвижение гражданских инициатив и освещение общественных проблем › Развитие сервисов в концепции «незаметного государства» › Подтверждение юридически значимых действий, фактов о себе, операций

Государство работает с данными всех типов (см. подробнее в главе 2). Госорганизации в качестве дополнительного источника аналитики используют в том числе данные, приобретенные у банков⁶², мобильных операторов связи, телеком-провайдеров, бюро кредитных историй. В них, как правило, не содержатся персональные данные (ПДн) отдельных граждан. Они используются при принятии локальных решений, например для мониторинга качества жизни населения и миграционных потоков, оценки транспортных потоков и фактической численности жителей конкретной территории, определения реального уровня доходов жителей региона, уровня социального неравенства в регионе, уровня кредитной нагрузки и т. п.

Одна из главных задач управления данными на уровне государства — обеспечение межведомственного взаимодействия, которое требует доступности данных онлайн на высокой скорости для всех исполнителей и получателей государственных услуг и сервисов: для ведомств, граждан, представителей бизнеса, других организаций. Всем клиентам государства хотелось бы получать его услуги так же легко, как это происходит сейчас, например, в банковской сфере: быстро, онлайн, без многократного ввода информации, уже имеющейся у поставщика услуг (банка или государства). Хранение бумажных документов постепенно перестает быть актуальным, и необходима альтернатива.



«Как гражданин видит целевое состояние межведомственного взаимодействия при оказании госуслуг? Гражданин обращается за услугой по любому удобному для него каналу (МФЦ, ЕПГУ, через региональные порталы, личные кабинеты ведомств) и проходит идентификацию привычным для него способом (паспорт, биометрия, логин/пароль ЕСИА или цифровая подпись). Он ожидает, что информация о нем, имеющаяся в государственных реестрах и системах, будет получена автоматически и ему не придется собирать „багаж“ документов, а затем приносить их в бумажном виде или загружать сканы на портал „Госуслуги“. Для достижения такого целевого состояния потребуются значительные изменения в процессах управления данными в ИС ведомств и процессах межведомственного обмена данными».

Олег Виниченко, исполнительный директор дирекции «Цифровые решения для государственного сектора» ПАО «Сбербанк»

Последние годы развивается ряд инициатив в сфере работы с госданными. В 2020–2021 годах в большинстве федеральных и региональных органов власти появились руководители ЦТ (Chief Digital Transformation Officers,

⁶² Например, в 2016 году Сбербанк запустил проект «Открытые данные» на основе больших данных, который с тех пор развился в отдельное направление «СберИндекс». URL: <https://sberindex.ru/ru>

CDTO), ответственные за проведение ЦТ в своих ведомствах, и в некоторых из органов власти уже есть ответственные за работу с данными (CDO). (Подробнее об функциях CDTO и CDO в государственной организации см. раздел 5.1.) В сфере оказания госуслуг акцент сместился с МФЦ на развитие Единого портала государственных и муниципальных услуг (ЕПГУ, портал «Госуслуги», gosuslugi.ru), проактивное оказание услуг, создание инструментов для сбора обратной связи по ним⁶³. В регионах создаются платформы обратной связи и центры управления регионом (ЦУР).

Система межведомственного электронного взаимодействия (СМЭВ) создавалась 10 лет назад для поддержки оказания госуслуг и до сих пор развивается, чтобы соответствовать современным запросам к скорости и функциональности работы. СМЭВ развивается в сторону межведомственного обмена теми блоками данных, которые нужны ведомствам для выполнения их функций, оказания госуслуг и реализации полномочий⁶⁴.

В числе приоритетных направлений деятельности Минцифры — перевод востребованных госуслуг в электронный формат и развитие национальной системы управления данными (НСУД).

Авторы раздела:



Д. А. Кашко



С. В. Сергиенко

3.1.2 НАЦИОНАЛЬНАЯ СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ

НСУД разрабатывалась⁶⁵ для систематизации работы с данными на государственном уровне и в основном регулирует сферу межведомственного взаимодействия. Задача НСУД — обеспечить согласованность госданных, удобную работу с ними, их доступность для участников взаимодействия.

Система включает четыре основных компонента (рис. 16).

1. Управляющий компонент — федеральная государственная информационная система «Единая информационная платформа НСУД» (ФГИС «ЕИП НСУД», nsud.info.gov.ru/), где ведомства формируют и ведут модели своих данных, а оператор на этой основе ведет единую модель госданных («карту данных»).

2. Ведомственные витрины данных — компоненты ведомственных систем, на которых ведомства публикуют данные в соответствии с моделью, настроенной в ЕИП НСУД.

⁶³ Дмитрий Чернышенко: цифровая платформа «Госуслуги. Решаем вместе» будет запущена во всех субъектах страны // Правительство России. URL: <http://government.ru/news/41722/>

⁶⁴ В 2010 году в законе «Об организации предоставления государственных и муниципальных услуг» от 27.07.2010 № 210-ФЗ было введено положение, которое должно было коренным образом изменить ситуацию: «Органы, предоставляющие государственные услуги, <...> не вправе требовать от заявителя представления документов и информации, которые находятся в распоряжении органов, предоставляющих государственные и муниципальные услуги, либо подведомственных организаций». URL: http://www.consultant.ru/document/cons_doc_LAW_103023/

⁶⁵ Постановление Правительства РФ от 14.05.2021 № 733 «Об утверждении положения о федеральной государственной информационной системе „Единая информационная платформа Национальной системы управления данными“» // КонсультантПлюс. URL: http://www.consultant.ru/document/cons_doc_LAW_384215/ См. также законопроект с поправками в 149-ФЗ «Об информации, информационных технологиях и о защите информации» в части формирования национальной системы управления данными» // Официальный портал проектов нормативных правовых актов. URL: <https://regulation.gov.ru/projects#npa=115660>



Рисунок 16. Основные компоненты ФГИС «ЕИП НСУД»

3. Обработчик и маршрутизатор запросов к данным — компонент СМЭВ («СМЭВ 4»), который обеспечивает прямой доступ к данным⁶⁶ ведомственных витрин.

4. Аналитическое хранилище — подсистема информационно-аналитического обеспечения ЕИП НСУД, отвечающая за сбор и аналитическую обработку данных, получаемых из витрин и других источников по поручениям органов, уполномоченных правительством РФ.

НСУД позволяет решать актуальные задачи управления госданными в современной дата-ориентированной экономике:

- › предоставление данных в режиме реального времени⁶⁷: обеспечение быстрого доступа к данным на витринах для онлайн-взаимодействия при оказании госуслуг и выполнении государственных функций;
- › возможность получать данные в режиме запроса либо в режиме подписки;
- › обеспечение доступности данных: возможность опубликовать новые наборы данных, найти опубликованные наборы и получить право доступа к ним в минимальный срок;

⁶⁶ В отличие от ориентированного на обмен электронными сообщениями «СМЭВ 3».

⁶⁷ Сейчас под режимом реального времени понимается такой режим взаимодействия ведомств, при котором с момента отправления межведомственного запроса до момента получения ответа на этот запрос проходит не более двух секунд.

- › контроль качества данных, включая их полноту, актуальность, непротиворечивость и связность (в том числе с эталонными данными);
- › построение единой модели данных (понимание, где какие данные находятся и в каком виде);
- › обеспечение возможности оперативной передачи данных в аналитическое хранилище и получения аналитических показателей.

Для представления данных с использованием НСУД обладателю данных (поставщику) необходимо:

- › описать свои информационные ресурсы и ИС в ЕИП НСУД, как минимум в части, подлежащей передаче (межведомственному обмену), создать и согласовать модель государственных данных;
- › создать витрину данных и обеспечить регулярное размещение в ней актуальных данных, соответствующих согласованной модели госданных;
- › сформировать и зарегистрировать регламентированные запросы к данным, размещенным на витрине, по которым допускается получение данных с витрины;
- › зарегистрировать витрину данных и регламентированные запросы в СМЭВ.

Для получения доступа к данным с использованием НСУД пользователю данных (получателю) необходимо:

- › найти нужные данные, пользуясь навигацией и поиском ЕИП НСУД;
- › получить доступ к соответствующим регламентированным запросам в личном кабинете СМЭВ;
- › настроить обработку получаемых из агента СМЭВ данных в своей ИС либо в витрине данных;
- › после получения разрешения на доступ к данным — выполнять необходимые запросы к данным либо получать данные по подписке.

Технологии НСУД позволяют максимально сократить процессы запроса и публикации необходимых данных. Создание нового регламентированного запроса состоит в написании и проверке SQL-запроса и занимает несколько часов. Публикация новых данных на витрине занимает не более нескольких дней и состоит из создания структуры данных, настройки ETL-инструментария (см. раздел 6.2) и реализации правил проверки качества.

Типовое ПО витрины данных⁶⁸ реализовано как свободное ПО и доступно в Национальном фонде алгоритмов и программ (portal.eskigov.ru/nfap/). Для публикации небольших наборов данных, не предъявляющих высоких требований к производительности, есть облегченный вариант витрины данных; он содержит только реляционную систему управления базами данных (СУБД, см. раздел 6.2) и доступен в виде исполняемых файлов.

⁶⁸ Витрины данных реализуют подход мультимодельности (он же Polyglot Persistence, многовариантное хранение), при котором данные хранятся одновременно в нескольких базах данных и ответ возвращается из той базы данных, которая способна быстрее других обработать запрос.

3.1.3 ОТКУДА ПОЯВЛЯЮТСЯ И ГДЕ ХРАНЯТСЯ ГОСДАННЫЕ

Авторы раздела:



О. А. Виниченко



А. А. Малахов



О. Б. Саваровская

Источниками госданных выступают буквально все граждане и организации, чьи данные тем или иным образом попадают в организации госсектора. Организации, граждане и должностные лица собирают и сдают всевозможную отчетность, а также создают данные в результате своей хозяйственной или управленческой деятельности; госорганизации ведут различные реестры и ИС для выполнения своих основных функций (обычно возникновение нового полномочия у госорганизации влечет за собой обязанность вести очередной реестр или ИС). Государство ведет себя так же, как любая крупная компания, — стремится собрать максимум данных, которые могут пригодиться для его задач.

Госданные хранятся не только в государственных, муниципальных и ведомственных ИС (ГИС — наиболее важные ресурсы), но также в отдельных электронных таблицах и базах (например, Excel) и даже на бумаге⁶⁹ — и всеми ими необходимо правильно управлять. Свои ИС имеют не только ФОИВ, но и государственные и муниципальные организации, подведомственные организации, казенные учреждения, унитарные предприятия и т. д. О многих ИС знают только те, кто непосредственно работает с ними, например об ИС в отдельных бюджетных учреждениях, созданных для нужд этих учреждений.

В России пока нет единого каталога госданных (он создается в ЕИП НСУД), зато есть около 600 различных реестров ИС, содержащих госданные. В этих реестрах — десятки тысяч ИС различных органов, ведомств, учреждений (о проблемах работы с данными в госсекторе см. раздел 3.2.1). В одной только ФГИС Координации информатизации (ФГИС КИ, portal.eskigov.ru), которая задумана и реализована как основной реестр ИС госорганизаций, зарегистрировано более 800 федеральных и 3300 региональных ИС⁷⁰.

Разобщенность госданных, существующих во множестве ИС, обусловлена сложностью федеративного устройства России и большим количеством ведомств и организаций разного уровня (государственных, муниципальных, бюджетных, подведомственных), финансируемых из бюджета. В госсекторе действуют сотни тысяч юридических лиц, и все они вовлечены в оборот данных: создают их, собирают, обрабатывают, хранят, передают и т. д.

⁶⁹ В России огромный объем информации хранится в многочисленных архивах в бумажном виде или в устаревших электронных форматах. Здания архивов территориально распределены по всей стране, что дополнительно затрудняет доступ к данным в них. Содержащиеся в этих документах «ретроспективные» данные (данные прошлых лет) еще предстоит оцифровать (или перевести в современные форматы) и включить в реестры ИС.

⁷⁰ Заместитель председателя Правительства РФ Д. Н. Чернышенко привел следующие данные: в России 826 федеральных и 3303 региональных информационных систем. См.: Конференция «Использование единой платформы „ГосТех“» 23.10.2021 // ГосТех. URL: <https://platform.digital.gov.ru/events/23102021/>



В России имеется около 20 000 органов местного самоуправления (по одному на каждое муниципальное образование)⁷¹, а также 90 ФОИВ⁷² и около 2500 РОИВ. По данным за 2014 год⁷³ только государственных и муниципальных учреждений в стране насчитывалось 178 875. В совокупности с организациями, финансируемыми из бюджета, это дает оценку в сотни тысяч юрлиц, приведенную выше.

Есть и другие факторы, затрудняющие управление государственными данными и их использование:

- › особенность госданных в реестрах — юридическая значимость записей, необходимых для выполнения государством его функций;
- › огромное количество данных разной значимости и детализации, собранных в разное время и с разными целями;
- › закрепление в многочисленных НПА конкретных практик и особенностей работы с госданными, создающее правовые барьеры;
- › уровень цифровой зрелости государства в работе с данными⁷⁴.

На практике в любой сфере первыми улучшаются наиболее востребованные данные и процессы. Так, с 2010 года после принятия ФЗ-210 активно развивается сфера оказания госуслуг в электронном виде⁷⁵. Достигнуты существенные результаты, которые стали особенно заметны во время пандемии⁷⁶. Процессы типизации и сквозного обмена постепенно налаживаются, получить госуслуги стало проще, их оказывают быстрее.



Сейчас на «вершине айсберга» находятся около 20 наиболее развитых ключевых информационных ресурсов, постоянно задействованных в межведомственном обмене. Они содержат информацию о юридических, физических лицах, имуществе, земельных ресурсах, финансово-бюджетной сфере (СНИЛС, ЕГРЮЛ, ЕГРИП, данные ПФР и т. д.).

Однако вне действия ФЗ-210 по-прежнему заметно отсутствие последовательной, ясной стратегии управления госданными на уровне государства, а у ведомств нет полномочий для изменения ситуации вне их компетенции.

⁷¹ См. доклад Минюста «Развитие федеративных отношений и местного самоуправления». URL: <https://minjust.gov.ru/ru/activity/directions/977/>

⁷² См. список на Сервере органов государственной власти Российской Федерации. URL: <http://www.gov.ru/main/ministry/isp-vlast44.html>

⁷³ См.: Кривко Т. С. Структура государственных (муниципальных) учреждений // Международный научно-исследовательский журнал. № 3 (34). Ч. 3. 2015. URL: <https://research-journal.org/economical/struktura-gosudarstvennyh-municipalnyh-uchrezhdenij/>

⁷⁴ Исследовательская и консалтинговая компания Gartner еще в 1995 году разработала модель «цикл хайпа» для оценки уровня зрелости технологий. Ситуацию с госданными в России сейчас можно отнести ко второму этапу, тогда как первым была хаотичная и рассогласованная автоматизация многочисленных государственных функций. См.: Understanding Gartner's Hype Cycles // Gartner. URL: <https://www.gartner.com/en/documents/3887767>

⁷⁵ Федеральный закон от 27.07.2010 № 210-ФЗ «Об организации предоставления государственных и муниципальных услуг».

⁷⁶ В 2020 году 56 млн человек воспользовались сервисами единого портала Госуслуг // Минцифры РФ. URL: <https://digital.gov.ru/ru/events/40942/>

Попытки развивать проактивное оказание госуслуг вскрывают проблемы с текущим состоянием данных в разных органах власти, на разных уровнях. Распространенная в госсекторе ситуация выглядит так:

- для решения конкретной задачи ведомству нужны данные, которые находятся в разных источниках и управляются разными операторами;
- многие необходимые данные, скорее всего, хранятся в неудобном формате или находятся на бумажном носителе, доступ к ним ограничен нормативно или организационно, а в некоторых случаях и вовсе отсутствует (подробнее см. раздел 3.2);
- бывает сложно выяснить, где найти необходимые данные, кто ими управляет, как получить к ним доступ, на каких условиях ими можно распорядиться.

Открытые государственные данные образуют отдельную категорию. Они должны быть доступны общественным (некоммерческим), коммерческим, исследовательским и образовательным организациям, СМИ⁷⁷. Открытые данные размещаются в интернете в виде систематизированных данных, организованных в формате, допускающем автоматическую обработку (без предварительного изменения человеком), в целях неоднократного, свободного и бесплатного использования⁷⁸. В конечном счете они касаются всех граждан, хотя при раскрытии данных важно соблюдать баланс между открытостью и приватностью⁷⁹.



Удачный пример государственного сервиса открытых данных — портал госзакупок zakupki.gov.ru (см. кейс ЕАЭС в разделе 3.2.2). В любой момент из ИС с помощью API можно скачать для анализа фрагмент базы данных, относящийся к определенной территории или тематике. Немаловажный фактор развития сервиса — спрос на раскрытие этих данных и последовательная политика государства по их раскрытию.

Государственные организации и ведомства — тоже потребители открытых данных; благодаря открытым данным они могут выяснять последствия тех или иных решений, находить взаимосвязи между ними, анализировать текущее положение дел и решать, что делать дальше. Нередко для самих госорганизаций пользоваться открытыми данными проще, чем получать данные по межведомственным каналам.

⁷⁷ Тему открытости данных в России много лет разрабатывает АНО «Инфокультура»; см., в частности, проект «Каталог данных», URL: <https://www.datacatalogs.ru> Рекомендуем экспертные доклады об открытости данных в ФОИВ: «Открытость государства в России» — 2021 // Счетная палата РФ, АНО «ЦПУР», АНО «Информационная культура», 2021. URL: <https://ach.gov.ru/upload/pdf/Otkrytost-2021.pdf>; Акиров О. И., Соловьев Д. Б. Открытые данные в дипломатии: Россия в контексте международного опыта / под ред. М. О. Комина. М.: ЦПУР, 2021. URL: https://cpur.ru/research_pdf/PP_opendata_in_diplomacy-2021.pdf

⁷⁸ См.: Методические рекомендации по публикации открытых данных государственными органами и органами местного самоуправления, а также технические требования к публикации открытых данных. Версия 3.0 (утв. протоколом заседания Правительственной комиссии по координации деятельности Открытого Правительства от 29.05.2014 № 4).

⁷⁹ Подробнее об этом см.: Вяткин И. О., Колин М. О., Копыток В. К. Модели институционализации доступа исследователей к данным государства: аналитический обзор // Центр перспективных управленческих решений. М.: ЦПУР, 2021. URL: https://cpur.ru/research_pdf/PP_data_access_models_2021_FIN_red.pdf



О. А. Виниченко



С. В. Коршунова



А. А. Малахов



О. Б. Саваровская

3.2 ТРУДНОСТИ В РАБОТЕ С ГОСУДАРСТВЕННЫМИ ДАННЫМИ

 **Время чтения: 40 мин.**

3.2.1 ОСНОВНЫЕ ТИПЫ ПРОБЛЕМ

Ниже описан в общем виде ряд проблем, которые по мнению экспертов, участвовавших в подготовке настоящего навигатора, оказывают критическое влияние на развитие управления данными в России.

Не претендуя на исчерпывающее описание проблем, их причин и сложных взаимосвязей, отметим, что эти проблемы проявляются и требуют решения одновременно на разных уровнях: внутри- и межведомственном, внешнем (предоставление госданных получателям не из госсектора). Минцифры сейчас предпринимает существенные действия, чтобы добиться системных изменений.

1. Низкое качество данных

Качественные данные (см. главу 7) — это фундамент использования данных в любых целях. У проблемы низкого качества госданных есть много проявлений:

- › нет системных механизмов обеспечения качества госданных, в каждом конкретном случае вопрос качества решается исходя из локальных приоритетов;
- › отсутствуют требования к работе с госданными, направленные на исключение возникновения ошибок;
- › нет общих стандартов, единых справочников, общих метаданных;
- › отсутствует мониторинг качества госданных и управление инцидентами, связанными с их качеством;
- › отсутствуют механизмы совместимости госданных при обмене, разрешения противоречий между госданными разных информационных ресурсов, приведения госданных в соответствие друг с другом;
- › госданные в «цифре» могут противоречить сведениям на бумаге; если приоритет отдается «бумажным» сведениям, снижается эффективность использования данных в электронном виде;
- › из-за низкого качества данных невозможно использовать их для оценки текущей ситуации.

Отсутствие единого подхода к исправлению ошибок в госданных критично при переходе на реестровую модель предоставления госуслуг (например, при выдаче разрешений). Особенность реестровой модели в том, что юридически значимыми данными в них являются данные реестров (записи

в реестрах), а не исходные документы⁸⁰. До сих пор при обнаружении ошибки следовало удалить запись и создать новую запись (или новый документ). В реестровой модели этот принцип не работает, так как в момент появления записи в реестре возникает юридически значимый факт. Необходимо разрабатывать новые правила и порядки, предлагать решения для исправления как технических ошибок, так и ошибок в сути, в частности описывать ситуации, когда к исправлению записи необходимо привлечь субъекта данных или поставщика данных.

Нельзя обеспечить взаимодействие между системами разных ведомств без потери смысла, не имея системы классификации и кодирования, единых справочников. В законодательстве описаны полномочия и функции ведомств: косвенным образом это описание предопределяет, кто какие системы ведет. Необходимы мастер-данные, эталонные источники и все, что поможет использовать достоверные качественные источники информации для автоматизации внутри ведомств или организаций, работающих с госданными, и межведомственного обмена.



Специфика каждого ведомства, прежде всего отраслевая, предполагает разную трактовку одних и тех же терминов. Возможен и противоположный вариант — разные названия в разных ведомствах для одних и тех же сущностей. Например, разный понятийный аппарат Минфина и Федеральной налоговой службы обусловлен Бюджетным кодексом (для Минфина) и Налоговым кодексом (для ФНС).

Единство нормативной справочной информации особенно важно при работе с массивами неструктурированных данных. Справочники могут быть национальными и отраслевыми и, соответственно, в обязательном порядке использоваться всеми участниками или на национальном уровне, или на отраслевом. В примере ниже справочники полов — отраслевые.



Что такое «пол»? В справочниках правоохранительной системы — пять полов, они определены исходя из того, что можно использовать при дознании (вторичные половые признаки, базовые и с отклонениями). В авиации «полов» тоже пять, но уже других: мужчина, женщина, «большой мужчина», «большая женщина» и «ребенок». В системе здравоохранения полов шесть — те, которые имеют медицинское значение, в том числе скрытые факторы пола.

Для обеспечения интероперабельности (см. раздел 7.1) необходимы единые требования к управлению госданными. Без таких требований

⁸⁰ См. исследование Аналитического центра при Правительстве РФ: Основные результаты анализа законодательства зарубежных государств и Российской Федерации в области управления данными, описания моделей правового регулирования управления данными. М., 2019. URL: https://ac.gov.ru/uploads/_Projects/PDF/NSUD/NSUD_regulation.pdf

не только нет ответственных за организацию и управление данными, но и отсутствует системный подход к обеспечению качества данных⁸¹.



В инфраструктуре электронного правительства действует Единая информационная система нормативной справочной информации (ГИС ЕИС НСИ). У этой системы есть методические и организационные недостатки. Например, непонятно, какой из десятка справочников ОКАТО/ОКТМО эталонный, какие из них можно использовать, а какие нельзя. ФОИВ, регионы, другие участники системы госуправления самостоятельно решают эту проблему, находя и встраивая в свои ИС подходящие справочники. Однако такой подход не обеспечивает интероперабельность: в разных справочниках содержание одного и того же термина может различаться.

На качество данных в ИС влияет еще и отсутствие форматно-логического контроля — проверки данных по критериям полноты, качества, формата и безопасности при вводе их пользователем. Там, где данные при вводе не проверяются по эталонным справочникам, можно быть уверенным в наличии «мусора» в данных.



В отсутствие эталонных справочников даже верные (по смыслу) данные могут заноситься вручную по-разному. Типичная картина — в датасете одного ведомства можно найти несколько вариантов наименований городов и регионов. Например, для Петрозаводска:

- › Городской округ г. Петрозаводск
- › Петрозаводский городской округ
- › город Петрозаводский
- › г. Петрозаводск
- › ГО Петрозаводск

2. Госорганизации вынуждены многократно собирать данные вместо получения их из эталонных источников

Десятки госорганизаций и ведомств могут собирать одни и те же данные для реализации своих функций, не синхронизируя их между собой или с эталонным источником (если он есть). Часто при этом неясно, какие данные эталонные, на записи какой системы стоит ориентироваться.

В качестве примера можно взять знакомую каждому ситуацию: ведомства, собирая наши данные, нередко дублируют их, многократно запрашивают одни и те же данные. На рис. 17 приведены примеры данных, которые собирают несколько крупнейших ведомств.

⁸¹ Сейчас в этом направлении предпринимаются некоторые шаги, но пока каждой организации приходится придерживаться собственных правил.

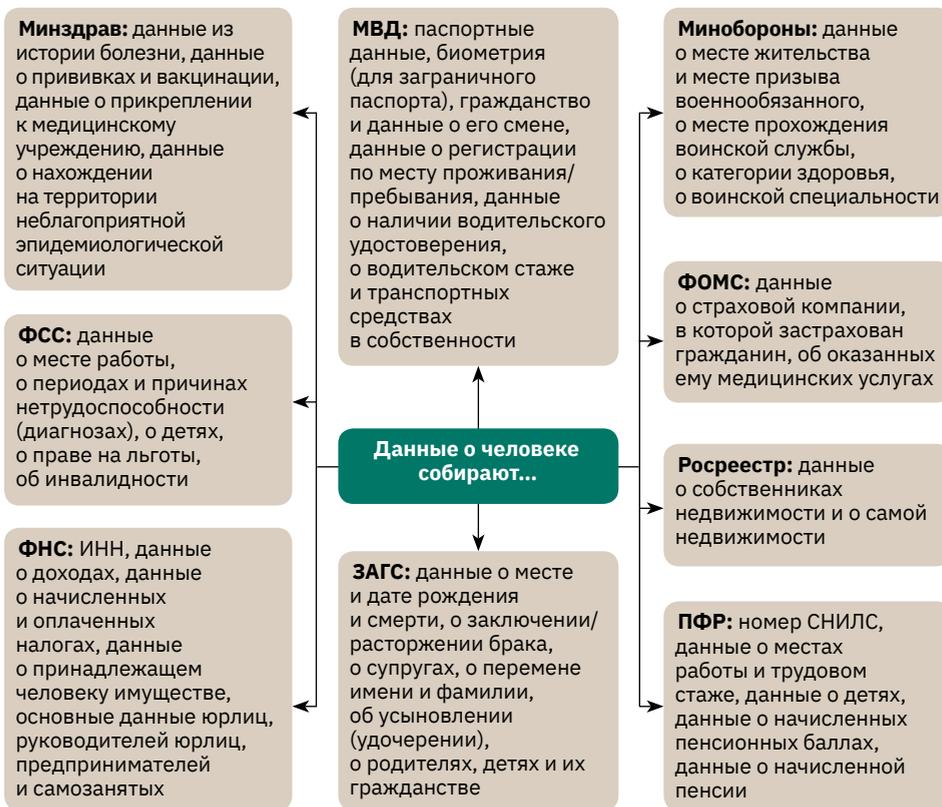


Рисунок 17. Данные о гражданах, которые собирают и присваивают крупнейшие госорганы

Дублирование ведет к увеличению расходов на госуправление и к сложности управления. В числе других проблем:

- составы данных не описаны полностью в одном документе (в ходе разработки ИС требования многократно уточняются, исправляются, обновляются в соответствии с новыми поручениями, изменения утверждаются разными актами);
- иногда дублировать сбор данных приходится из-за правовой недоступности уже собранных данных;
- без единого каталога госданных и организаций, ответственных за их ведение (создание), о некоторых данных, полученных в ходе информационного обмена, знают только стороны такого обмена;
- организациям приходится многократно сдавать одни и те же данные в виде отчетности в разных форматах, в разные сроки и в разные ведомства, что создает излишнюю нагрузку на всех участников процесса.

Сейчас получателям данных — как государственным структурам, так и бизнесу — приходится самостоятельно (и под свою ответственность) решать, какая информация верна и актуальна.

3. Сложно получить госданные

Любая госструктура, как правило, выступает как монополист, когда речь идет о сборе и обработке тех уникальных данных, которые находятся в ее ведении, или о распоряжении ими⁸². В понятие «информационная система» входят содержащиеся в ней данные⁸³; эксплуатацией ИС и данных в ней занимается оператор ИС. На практике наполнением ИС и распоряжением данными в ней занимаются разные функциональные структуры оператора ИС, а соответствующие полномочия нередко оформлены локальными актами. В отсутствие четкого разграничения предмета управления и функций возникают «белые пятна» в управлении и размываются границы ответственности; в результате госорганизации часто действуют по принципу «запрещено все, что не разрешено».



Например, госорганизация — оператор ИС отказывается предоставить коммерческим организациям данные из этой ИС. Оператор не имеет права предоставлять первичные данные (документы) третьим лицам (это в том числе обеспечивает защиту данных), а чтобы делиться агрегированными данными, следовало предусмотреть такую возможность на этапе разработки ИС.

Из-за разобщенности и наличия тысяч ИС потребителям госданных (о потребителях госданных и их целях см. раздел 3.1.1) сложно найти данные, определить права и ответственность по отношению к данным, выяснить, где и как происходит работа с данными.



«Многие исследователи, уже работающие с государственными данными, отмечают сложность получения доступа к необходимой информации, неудобный формат, отсутствие системы и методологии в подаче»⁸⁴.

**Мария Шклярчук, академический директор
Центра подготовки РКЦТ**

Получить доступ к госданным может быть сложно по разным причинам:

- › не определены или нечетко определены права доступа к госданным (что связано, в частности, с данными ограниченного распространения);

⁸² Отраслевые и ведомственные «колодцы» затрудняют обмен данными. Одни ведомства в соответствии со своими полномочиями создают и собирают уникальные данные (МВД, ЗАГС, ФНС и другие), другие ведомства дублируют эти данные, когда собирают их для осуществления собственных функций.

⁸³ Информационная система — совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств. Оператор ИС — гражданин или юридическое лицо, осуществляющие деятельность по эксплуатации ИС, в том числе по обработке информации, содержащейся в ее базах данных. См. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».

⁸⁴ Сапрыкина А., Мельникова Ю. Ученым облегчат работу с госданными // ComNews. URL: <https://www.comnews.ru/content/209399/2020-10-05/2020-w41/uchenym-oblegchat-rabotu-gosdannymi>

- › некоторые данные невозможно получить быстро и онлайн, хотя формально они доступны без ограничений;
- › нет единого подхода к платному/бесплатному предоставлению госданных;
- › данные в ИС существуют в неудобных форматах (например, сканы в формате pdf могут называться машиночитаемыми данными);
- › необходимые данные вообще не существуют или не собираются;
- › не решен вопрос о допустимом использовании некоторых данных (в частности, ПДн или больших данных, необходимых для обучения ИИ) с соблюдением баланса интересов всех сторон.

Неконсистентность способов взаимодействия между госструктурами существенно осложняет регулирование доступа госорганов к данным. Во многих случаях неизвестно:

- › какие государственные структуры какими данными управляют (кто владелец данных);
- › на каком основании они это делают;
- › какие права есть у ведомств на использование данных других ведомств.

Госорганизации на всех уровнях власти сталкиваются с тем, что предоставление данных из ГИС регламентировано, и поставщики данных определенного реестра не всегда могут выступать получателями данных из этого же реестра⁸⁵. Это ведет к повторному сбору информации (см. проблему 2).

В России более 50 «тайн» — видов информации ограниченного доступа (государственная тайна, служебная, врачебная, налоговая, ПДн и др.). Ограничения доступа к этим данным в ГИС и других ИС для третьих лиц регулируются рядом НПА. При этом требования по обезличиванию данных в соответствии с видами тайн не определены однозначно и полностью (подробнее о правовых аспектах работы с данными см. приложение В).



Несовершенство НПА называют в качестве трудности при работе с данными 44% всех опрошенных и 67% руководителей высшего звена (из числа опрошенных)⁸⁶.

Проблема «лоскутного» законодательства в сфере данных⁸⁷ существует не только в России. В странах ЕС континентальной правовой семьи (Франции,

⁸⁵ Например, в ГАС «Управление» есть более 50 ролей с разным уровнем доступа. См.: Регламент подключения и интеграции в ГАС «Управление». URL: <https://gasu.gov.ru/preview?fileId=213>

⁸⁶ По результатам опроса, проведенного Центром подготовки РКЦТ летом 2021 года. См. также введение.

⁸⁷ Подробнее о законодательстве разных стран в сфере работы с данными см. исследование Аналитического центра при Правительстве РФ: Основные результаты анализа законодательства зарубежных государств и Российской Федерации в области управления данными, описания моделей правового регулирования управления данными. М., 2019. URL: https://ac.gov.ru/uploads/_Projects/PDF/NSUD/NSUD_regulation.pdf

Германии, Эстонии) правовое регулирование управления данными разделяет законодательство в сфере обеспечения доступа к информации публичного сектора и регулирование в сфере защиты ПДн. Действуют также отдельные законы о разных видах информации ограниченного доступа (о государственной тайне, о коммерческой тайне и др.).



В ФЗ-143⁸⁸ о Едином госреестре записей актов гражданского состояния (ЕГР ЗАГС) был указан закрытый перечень лиц, имеющих право получать справки о смерти, и Минцифры в этот перечень не входило. В результате Минцифры не имело юридических оснований, чтобы удалять аккаунты умерших людей из базы ЕСИА. Внесение изменений на уровне федерального закона заняло полтора года.

Ситуация с данными, обладающими юридической значимостью, медленно улучшается. Ситуация с открытыми данными в России скорее ухудшается: публикуется мало новых наборов данных, отменяются обязательства по публикации некоторых данных в открытом формате (данные «закрываются»⁸⁹), а уже опубликованные наборы не актуализируются и постепенно теряют ценность, поскольку содержат неверные или устаревшие сведения. У потребителей открытых данных мало или нет средств для обратной связи и диалога с разработчиками, а у разработчиков не хватает мотивации для повышения качества невостребованных данных.

4. Не проработаны вопросы обращения с ПДн и доступа граждан к данным о себе

Пока не разрешено противоречие, касающееся предоставления данных. С одной стороны, государство обязано защищать ПДн граждан и ряд других видов данных, и этому помогают строгое регулирование и сложный доступ. С другой стороны, для развития цифровой экономики как можно больше госданных должны быть доступны всем заинтересованным сторонам. Кроме того, гражданам полезно иметь доступ ко всем юридически значимым записям о себе, иметь возможность исправить ошибку, удалить (обновить) устаревшие данные, что сейчас практически невозможно.

В области этичного обращения с данными (см. также заключение) в государстве до сих пор существует ряд проблем и противоречий:

- › данные собираются избыточно;
- › данные собираются на чрезмерно длительный срок (срок хранения данных на «Госуслугах» — 50 лет);
- › данные, собранные для конкретной цели, иногда имеют необоснованно длительный срок хранения и не удаляются вовремя;

⁸⁸ Федеральный закон от 15.11.1997 № 143-ФЗ «Об актах гражданского состояния».

⁸⁹ В Кремле объяснили необходимость засекретить госзакупки Росгвардии и ФСО // Москва24. URL: <https://www.m24.ru/news/vlast/03112021/190506>

- › не до конца проработаны юридические механизмы, связанные с анонимизацией и деперсонификацией данных;
- › нет точного определения данных ограниченного распространения, неясно, как и кто может их обрабатывать.



Соглашаясь на обработку персональных данных при оформлении цифрового пропуска весной 2020 года, москвичи передавали не только персональные данные, но и информацию о месте работы, данные документов о гражданстве, образовании, контактные данные, технические данные устройств и другие. В соглашении говорилось, что данные собирают сроком на 10 лет с правом на обработку третьими лицами, в том числе с такими целями, как получение материалов «информационного и (или) рекламного характера»⁹⁰.

Доступ госструктур к данным может быть существенно затруднен из-за закрытого списка получателей данных, особого назначения сбора данных и по другим причинам. Однако если удалось найти правовые основания для доступа к данным и согласовать такой доступ, появляется большой риск бесконтрольного неограниченного доступа к данным и их распространения без необходимого участия первоисточника.

Во многих случаях не существует порядка использования данных, полученных от другого ведомства. Если порядки защиты данных и обеспечения их безопасности подробно описаны для первичного сбора, то при передаче данных непонятно, как получатель может ими распоряжаться, например, может ли далее их передавать и в каком объеме (возможные методы защиты чувствительных данных, опробованные в Роструде, показаны⁹¹ на рис. 18).

Обычно назначение собираемых данных прямо указывается в НПА или в согласии на обработку ПДн, и тогда такие данные (в том числе предобработанные, например обезличенные/деперсонифицированные) сложно или невозможно использовать повторно с другими целями. Исключение было сделано для оказания госуслуг, и это решение работает довольно успешно. Время от времени меняются требования Роскомнадзора к работе с обезличенными и необезличенными ПДн, критерии признания данных персональными⁹², в том числе как следствие развития и доработки законодательства о защите ПДн.

Государство обладает уникальным составом данных, но далеко не в полной мере использует их потенциал как реального инструмента госуправления. Нынешний период можно назвать переходным, когда органы власти знают

⁹⁰ Коршунова С. В., Потапова Е. Г. Данные — удалить, слежку — остановить // Центр подготовки РКЦТ. URL: <https://cdto.ranepa.ru/deletedata>

⁹¹ Из выступления Д. А. Васильева, заместителя руководителя Федеральной службы по труду и занятости, на секции «Лучшие практики доказательного подхода в органах власти в России» конференции АСОПП-2021 19 октября 2021 года. URL: <https://www.youtube.com/watch?v=x4VAxI96Wjk>

⁹² Аржанова Я. Как ужесточились требования к работе с персональными данными в 2021 году // Контур. URL: <https://kontur.ru/articles/4816>



Рисунок 18. Методы защиты чувствительных данных в Роструде (план)

о проблемах и их причинах. В следующем разделе описаны причины и те шаги, которые госорганы (в первую очередь, Минцифры) предпринимают для решения этих проблем.

3.2.2 ПРИЧИНЫ ПРОБЛЕМ С ДАННЫМИ И ПУТИ ИХ РЕШЕНИЯ

Анализируя ниже наиболее значимые причины проблем с данными, авторы сознательно не стали соотносить конкретные проблемы с теми или иными причинами, чтобы не усложнять описание: у каждой проблемы могут быть несколько причин, в разной степени «ответственных» за нее.

1. Разрозненность ИС («зоопарк информационных систем», «лоскутная информатизация»)

При автоматизации госфункций и процессов на основе разных подходов к управлению данными были реализованы разные архитектурные решения. В результате для обеспечения межведомственного взаимодействия ведомства вынуждены заключать между собой соглашения об обмене данными, оговаривая в них конкретные технологии, унификацию обмена, технологическую поддержку обмена и контроля, защиту данных от утечек, предоставление данных при условии наличия полномочий.



Основным препятствием для внедрения СМЭВ был разнотребовательность нормативных требований к виду и содержанию информации в разных органах. Данные, на основе которых выписывали справки, хранились в бумажных архивах и ИС, часто они были избыточными или недостаточными, неverified, требовали длительных ручных операций для их извлечения и подготовки и т. п.,

а неправильно составленная справка была проблемой заявителя. С внедрением СМЭВ органы должны решать подобные проблемы между собой и в предельно сжатые сроки, так как по закону ответ на межведомственный запрос должен быть направлен в течение пяти рабочих дней⁹³. С 1 января 2022 года этот срок должен будет составлять не более 48 часов, а в отдельных случаях — две секунды (взаимодействие в режиме реального времени)⁹⁴.

Как следствие объемного и довольно жесткого законодательного регулирования ГИС⁹⁵ появляются ведомственные ИС (ВИС), требования к которым проще и легче выполнять. Это усугубляет разрозненность вместо ее преодоления. Множество не согласованных между собой ИС («лоскутное одеяло») затрудняет не только межведомственное взаимодействие, но и использование данных самим ведомством — оператором данных, и анализ данных, и качество решений, принятых на анализе таких данных.

Подходы и возможные решения. Как уже упоминалось выше, такая ситуация с разрозненностью данных закономерна, но понятны и способы ее преодоления. Российскому госсектору еще предстоит обеспечить связность разных данных в многочисленных ИС и стандартизировать работу с данными в госорганизациях. Проблему выбора эталонов можно решить, определив эталонные источники данных (что планируется сделать в НСУД), назначив ведомства, ответственные за каждый такой источник, и приняв дополнительные НПА.

Одно из главных направлений системной работы с данными на уровне государства и в нормативной, и в технологической плоскости — создание общегосударственной модели (карты) данных с единым понятийным аппаратом. Созданием модели данных и ведением реестра информационных ресурсов сейчас занимается Минцифры. Необходима инвентаризация всех ИС с госданными (от ведомственных до ГИС), кросс-валидация систем и проверка связности собираемых в них данных. Отдельно нужно провести аудит данных внутри ведомств и аудит межведомственных данных, начиная с пилотных проектов для отработки технологии.

По результатам такого глобального аудита можно будет сформировать общий каталог государственных данных (о каталоге как инструменте управления данными см. раздел 7.2) и их владельцев, выявить существующие ограничения в НПА на уровне ведомств и подготовить предложения об изменениях в НПА в своей отрасли.

⁹³ Ст. 7.2 ч. 3 Федерального закона от 27.07.2010 № 210-ФЗ «Об организации предоставления государственных и муниципальных услуг»

⁹⁴ Постановление Правительства РФ от 23.06.2021 № 963 «Об утверждении Правил межведомственного информационного взаимодействия при предоставлении государственных и муниципальных услуг, в том числе рекомендуемых правил организации межведомственного информационного взаимодействия между исполнительными органами государственной власти субъектов Российской Федерации и (или) органами местного самоуправления». URL: <http://static.government.ru/media/files/8t93ZN3FupkxG7moA6rA4dvBRqhoV390.pdf>

⁹⁵ Историю ГИС и ключевые проблемы регулирования ГИС в России подробнее см. в аналитическом докладе «Оценка открытости государственных информационных систем в России». АНО «ЦПУР», АНО «Информационная культура», 2020. URL: <http://audit.gov.ru/upload/pdf/Оценка%20открытости%20государственных%20информационных%20систем%20в%20России.pdf>

2. Внесение данных вручную

Сбор отчетности в архаичной форме (в «бумажной» форме, в таблицах и с ручным вводом) неудобен для организаций и граждан, при этом создает избыточную нагрузку на субъектов, предоставляющих отчетность, и оставляет массу возможностей для появления ошибок и искажений. Из-за длительного процесса сбора отчетности к моменту принятия решений результаты анализа могут быть уже устаревшими и ненужными.



Пандемия COVID-19 в 2020 году показала недоступность оперативных данных, в первую очередь из системы здравоохранения. Для принятия оперативных решений данные, поступающие с месячной задержкой, не годились. В качестве временной меры медработникам пришлось заполнять отчеты ежедневно вручную.

Качество решений, принимаемых на основе данных, напрямую зависит от качества и актуальности самих данных. В России, как и во всем мире, возник запрос на перестройку системы сбора статистики и аналитики в режим ежедневного получения актуальных данных (вместо привычной периодически собираемой отчетности с агрегированными данными).

Сильная агрегация понижает ценность данных. Например, из данных, агрегированных на уровне региона или федерации, невозможно получить детальную информацию о текущей ситуации на уровне муниципалитета или района города. С учетом уровня развития инструментов и методов анализа данных сейчас востребованы именно первичные данные с максимальной детализацией и возможностью обогащать их другими данными для получения новой информации.

Подходы и возможные решения. Агрегированные данные должны существовать не вместо первичных данных, а наряду с ними. Нужны оперативные данные, поступающие в онлайн-режиме, в том числе с ретроспективой.



Проблема «грязных» данных активно обсуждалась в начале внедрения СМЭВ. Директор департамента информационных технологий и связи Правительства РФ Алексей Попов в 2011 году говорил о том, что большой объем данных пришлось срочно вводить в ИС руками, что привело к росту числа ошибок⁹⁶. Эксперты предполагали, что проблема частично решится либо в процессе активного взаимодействия органов, либо при естественном устаревании «грязных» данных и выходе их из оборота⁹⁷.

⁹⁶ Система межведомственного электронного взаимодействия (СМЭВ) // TAdviser. URL: [https://www.tadviser.ru/index.php/Продукт:Система_межведомственного_электронного_взаимодействия_\(СМЭВ\)](https://www.tadviser.ru/index.php/Продукт:Система_межведомственного_электронного_взаимодействия_(СМЭВ))

⁹⁷ Мальцев С. Ведущий эксперт по управлению документацией ЭОС Наталья Храмцовская: «Даже если завтра СМЭВ будет закрыта и заменена чем-то иным, определенное положительное влияние она уже оказала» // ICT Online. URL: <https://ict-online.ru/interview/183346/>

Вместо медленного по современным меркам ручного сбора необходимо наладить процессы непрерывного автоматизированного сбора данных с использованием эталонных справочников, а также модернизировать (или создать) инфраструктуру для сбора данных и предоставления доступа к ним. В положения о ФОИВ придется внести положения о роли поставщика данных и необходимых полномочиях. Внутри ФОИВ необходимо сформировать рабочие процессы по созданию новых наборов данных как продуктов, постепенно увеличивая их (наборов) количество.

3. Ведомственные барьеры

Ведомства долгое время решали проблемы с данными своими силами, по собственным планам, без коммуникации и кооперации, что объясняется в том числе уже упомянутой разрозненностью ведомственных ИС.



Ведомственные и внутриорганизационные барьеры называют проблемой в работе с данными 44% респондентов, а среди руководителей проектов и руководителей высшего звена — более 70% респондентов⁹⁸.

Большинство органов власти работает с госданными внутри своих полузамкнутых «отраслевых колодцев» (что создает риск «феодализма данных», см. раздел 8.2), а Минцифры не хватает полномочий для организации единой системы работы с данными для всех органов власти.



Сейчас даже на ЕПГУ гражданин иногда вынужден самостоятельно запрашивать и передавать из одной государственной организации в другую информацию (справки, выписки и т. п.), которую при налаженном межведомственном взаимодействии организации быстрее и проще передали бы напрямую.

Подходы и возможные решения. Необходимо наладить обмен данными между ФОИВ, а ведомствам — сотрудничать друг с другом, чтобы обеспечивать связность, полноту и историчность данных (шаги к такой кооперации сейчас делаются на уровне отдельных ФОИВ). Предстоит выработать единую методологию сбора данных, с тем чтобы:

- › установить прозрачные правила предоставления доступа к данным (для этого, например, собрать набор практик или кодекс для работы с данными в федеральных органах власти);
- › определить универсальные подходы к сбору данных внутри госорганов, в частности к методикам сбора данных и их типам;
- › облегчить коммуникацию между ведомствами.

⁹⁸ По результатам опроса, проведенного Центром подготовки РКЦТ летом 2021 года. См. также введение.

Публикация данных в формате открытых данных упрощает доступ к ним не только внешним потребителям (например, бизнесу), но и государственным организациям. Для этого необходимо провести инвентаризацию и анализ нормативных документов, затрагивающих открытые данные⁹⁹, актуализировать НПА, предусмотреть добавление в ГИС возможностей API и обновить «Методические рекомендации»¹⁰⁰: отказаться от устаревших практик формирования и передачи данных и добавить требования к архивам данных.

4. Низкий уровень компетенций и нехватка кадров

У госзаказчика часто не хватает кадров, обладающих компетенциями в области управления данными и проектами с использованием данных. Низкий уровень оплаты труда и высокая зарегулированность — серьезные препятствия для привлечения в госсектор квалифицированных специалистов и топ-менеджеров сферы ИТ.

Нельзя нанять аналитика, CDO, целую команду и принимать результаты их труда некритично, на веру. Руководителю нужно быть компетентным заказчиком, чтобы опираться на данные при принятии решений: уметь задавать корректные вопросы к данным и результатам их анализа; понимать, стоит ли доверять ответам; отличать достоверную аналитику от искажений и последствий влияния человеческого фактора; иметь представление о возможных логических ошибках. Хорошо известна «настраиваемость» социологических опросов: можно получать разные результаты, если по-разному сформулировать вопросы и предлагать разные наборы вариантов ответов¹⁰¹.

Подходы и возможные решения. Следует выделить ресурсы на создание инфраструктуры для «производства» качественных данных внутри ведомства, на обучение сотрудников ведомств работе с данными, на привлечение в штат квалифицированных специалистов по работе с данными. Каждому ведомству нужна своя внутренняя команда по работе с данными — от CDO до стюардов данных (см. раздел 5.1). Предстоит определить ответственных за все аспекты качества и стадии обработки госданных с соответствующими полномочиями.

5. Собираемые данные не используются

Часть госданных не используется для поиска лучших решений, а лежит «мертвым грузом». Иногда организационная культура не позволяет госорганизации открыто продемонстрировать результаты своей работы

⁹⁹ В первую очередь речь идет о таких документах, как 1) распоряжение от 10.07.2013 № 1187-р «О перечнях общедоступной информации, размещаемой в сети Интернет в форме открытых данных»; 2) Федеральный закон от 29.11.2007 № 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации»; 3) Федеральный закон от 06.10.2003 № 131-ФЗ (ред. от 01.07.2021) «Об общих принципах организации местного самоуправления в Российской Федерации».

¹⁰⁰ Методические рекомендации по публикации открытых данных государственными органами и органами местного самоуправления, а также технические требования к публикации открытых данных. Версия 3.0 (утв. протоколом заседания Правительственной комиссии по координации деятельности Открытого Правительства от 29.05.2014 № 4).

¹⁰¹ См. например: Аверьянов Л. Я. Социология: искусство задавать вопросы. М., 1998.

и раскрывать данные, показывающие реальную ситуацию, за которую отвечает эта организация. При этом чем чаще и активнее используются данные, чем больше потребителей в них заинтересовано, тем больше внимания уделяется качеству данных и управлению данными.

Подходы и возможные решения. Госорганам, в частности Минцифры, предстоит провести масштабные исследования для изучения спроса на госданные среди целевых групп (см. раздел 3.1.1), причем результаты исследований должны стать публичными. Нужно расширять обмен лучшими практиками как между ведомствами, так и между целевыми группами потребителей госданных. Возможным решением станут советы (или рабочие группы) по данным (см. раздел 5.1) на уровне ведомств для непосредственного взаимодействия с практикующими специалистами.



ПРОБЛЕМЫ ОТКРЫТОСТИ ДАННЫХ В СТРАНАХ ЕАЭС

Для эффективного использования возможностей единого экономического пространства ЕАЭС важно обеспечить равный доступ к информации, необходимой для принятия решений и выполнения требований национального законодательства. Так, ФНС России требует от компаний осмотрительности при выборе контрагентов, правительство РФ обязывает госкорпорации проверять конечных собственников компаний-поставщиков, законодательство РФ ограничивает работу с офшорными компаниями. При этом если Россия обеспечивает высокую прозрачность бизнеса, в том числе за счет открытости и бесплатности федеральных реестров юридических лиц, то в других государствах ЕАЭС складывается иная ситуация (см. табл. 2).

Сейчас в ЕАЭС отсутствует единый стандарт раскрытия информации, необходимой компаниям для принятия бизнес-решений и уменьшения рисков. В частности, в Казахстане из всего массива сведений о юрлицах, хранящихся в государственной базе данных «Юридические лица», в машиночитаемом формате возможно получить только пять категорий данных (для сравнения: из российского реестра предоставляется более тридцати категорий). Причем среди этих категорий нет информации о владельцах компаний — ее можно получить, только заказав выписку из указанного реестра. В Белоруссии в машиночитаемом формате можно получить только сведения о факте регистрации компании.

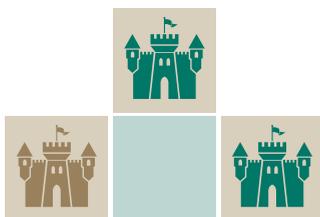
В России прозрачность госзакупок обеспечена многочисленными информационно-аналитическими системами, обслуживающими участников рынка госзаказа. Функционирование этих систем возможно благодаря предоставлению данных с официального сайта zakupki.gov.ru в машиночитаемом формате (формате открытых данных).

Необходимо унифицировать условия предоставления информации из реестров юридических лиц ЕАЭС, в том числе в формате открытых данных, а также определить единый набор сведений, которые должны раскрываться о юридических лицах¹⁰².

¹⁰² Кейс предоставлен О. В. Полетаевым, директором по развитию цифрового бизнеса группы «Интерфакс». Впервые кейс публично обсуждался на круглом столе «Экономика данных. Цифровое ускорение глобального сотрудничества» на конференции «Цифровая индустрия промышленной России» (ЦИПР) 29 июня 2021 года. URL: <https://eec.eaunion.org/news/soglashenie-ob-oborote-dannyih-eaes-sozdast-novye-usloviya-dlya-biznesa/> видео: <https://www.youtube.com/watch?v=t3TKPp6MCEM>

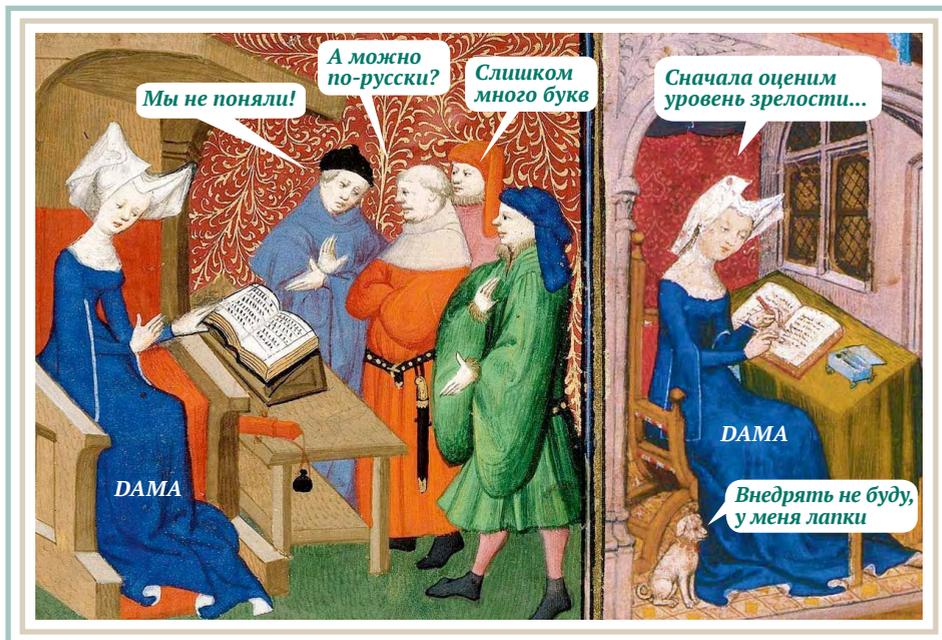
Таблица 2. Ситуация с открытостью данных о юридических лицах в странах ЕАЭС

Страна	Реестр регистрационных данных	Регистрация данных юрлиц от органов статистики	Бухгалтерская отчетность	Госзакупки	Банкротства
Россия	Оператор — ФНС. Данные по отдельным юрлицам предоставляются бесплатно на сайте. За плату доступен весь массив в машиночитаемом формате с регулярным обновлением	Данные предоставляются за плату	Бухотчетность за 2019–2020 годы размещается на спецпортале ФНС. Данные по отдельным юрлицам предоставляются бесплатно на сайте. За плату можно будет получать весь массив в машиночитаемом формате с регулярным обновлением. До 2019 года все данные размещались Росстатом бесплатно	Данные доступны на сайте госзакупок бесплатно, в том числе в машиночитаемом формате	Данные размещаются на портале ЕФРСБ, за минимальную плату доступен API, данные еженедельно публикуются в газете «Коммерсантъ»
Казахстан	Оператор — Минюст. Данные предоставляются бесплатно через сайт при наличии электронного сертификата, который может иметь только резидент Казахстана. Нерезиденты получают услугу платно после личного обращения в Минюст. Нет возможности выгрузки данных	Бюро национальной статистики размещает данные о предприятиях, но без необходимого набора полей. Бесплатно, данные для скачивания в машиночитаемом формате	Бухгалтерская отчетность только по компаниям публичного интереса, всего около 4 тыс. компаний. Данные доступны онлайн бесплатно и в виде массива за плату. Оператор — Депозитарий финансовой отчетности	Данные доступны на сайте госзакупок бесплатно, в том числе в машиночитаемом формате	Список банкротств обновляется раз в полгода на сайте Комитета государственных доходов Минфина РК
Белоруссия	Оператор — Минюст. Данные предоставляются за плату при личном обращении. На портале Единого госреестра онлайн можно узнать только факт регистрации и статус компании	Данные официально не предоставляются	Финансовые показатели доступны только по открытым АО (около 5 тыс. компаний). Бесплатно размещаются на Едином портале финансового рынка РБ. Данные в машиночитаемом формате отсутствуют	Данные доступны бесплатно на сайте госзакупок. Данные для скачивания отсутствуют	Бесплатно на сайте Единого государственного реестра сведений о банкротстве
Киргизия	Оператор — Минюст. Данные предоставляются бесплатно через сайт, есть машиночитаемый формат. Практически весь набор необходимых данных в наличии	Данные официально не предоставляются	Данные официально не предоставляются	Данные доступны на сайте госзакупок бесплатно, в том числе в машиночитаемом формате	Данные официально не предоставляются
Армения	Оператор — Агентство госрегистра Минюста Армении. Все данные только на армянском языке, в том числе машиночитаемые, доступны только резидентам Армении	Статистический портал открытых данных в стадии разработки	Планируется предоставлять данные по компаниям публичного интереса	Данные только на армянском языке, в машиночитаемом виде не предоставляются	Данные официально не предоставляются. Доступна информация только о конкурсных управляющих



БЛОК ІІ

НАЧИНАЕМ РАБОТАТЬ



4. СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ

 *Время чтения: 27 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Жизненный цикл данных начинается с планирования и заканчивается ликвидацией данных, он включает, к примеру, обеспечение доступности данных; хранение, обслуживание данных и предоставление совместного доступа к ним; повышение ценности данных.
- › Управление на основе данных невозможно без управления данными, включающего в себя, прежде всего, отлаженные процессы, сформированную архитектуру данных, инструменты обеспечения их данных.
- › Корпоративная модель данных включает наименования элементов данных, подробные определения данных и метаданных, концептуальные и логические сущности и связи между ними, а также бизнес-правила.
- › Работе с данными может помешать недостаток дата-центричного мышления, отсутствие квалифицированного руководства в области данных, несогласованность между планами в области данных и ИТ-проектами.
- › При разработке и внедрении системы управления данными приоритетными будут такие вопросы, как оценка зрелости существующей системы управления данными; разработка стратегии работы с данными и стратегии управления данными, тесно увязанных со стратегией организации; внедрение в организации функции руководства данными.

4.1 ЖИЗНЕННЫЙ ЦИКЛ ДАННЫХ В DAMA-DMBOK

Все управленческие решения и практические действия, имеющие отношение к данным, так или иначе привязаны к их жизненному циклу (ЖЦ). Он начинается с планирования и заканчивается ликвидацией данных. На рис. 19 представлены следующие процессы ЖЦ данных:

- 1) планирование;
- 2) проектирование и обеспечение доступности данных;
- 3) создание или получение данных;
- 4) перемещение, преобразование, хранение, а также обслуживание данных и предоставление совместного доступа к ним;
- 5) расширение возможностей использования данных;
- 6) улучшение (повышение ценности) данных.

Завершает цикл ликвидация данных. Кроме того, на протяжении всего ЖЦ данные можно очищать, преобразовывать, подвергать слиянию или агрегировать (см. раздел 6.2).

Требования к организации отдельных фаз ЖЦ зависят от вида данных (их классификацию см. в главе 2). ЖЦ данных в отдельно взятой организации может оказаться весьма запутанным. Наряду с ЖЦ данные имеют еще и путь (lineage), по которому они движутся от места возникновения до места использования; этот путь иногда называют также цепочкой данных.

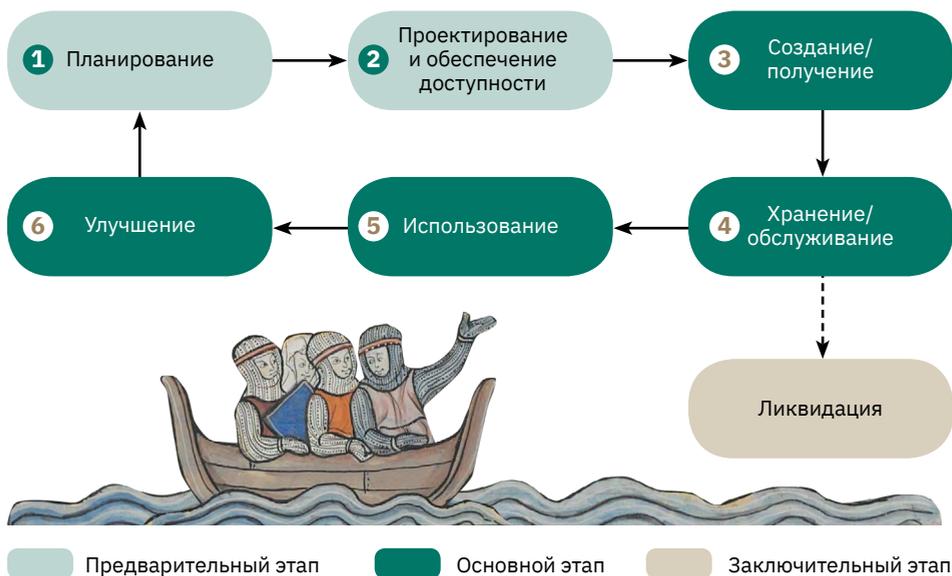


Рисунок 19. Жизненный цикл данных

Авторы главы:



А. С. Константинов



Н. В. Скворцов



Р. С. Стрекаловский

Следует иметь в виду, что перед госсектором стоит отдельная методологическая задача — адаптировать общие подходы DMBOK с учетом особенностей государственных организаций.

Схема на рис. 20 соотносит ЖЦ данных с основным фреймворком DAMA (см. раздел 1.3) и исходит из цели управления данными — дать организации возможность извлекать **выгоду из информационных активов**. Для этого требуется управлять жизненным циклом данных, поэтому функции, относящиеся к его конкретным фазам, помещены в центр диаграммы.



Рисунок 20. Фреймворк функций управления данными (в привязке к жизненному циклу)

Начинается все с **планирования и проектирования** надежных и качественных данных. К этому этапу относятся процедуры создания архитектуры и модели данных, разработка системы метаданных.



Архитектура данных образует фундамент управления данными. Цель архитектуры данных — служить мостом между бизнес-стратегией и ее технологической реализацией. Поскольку весь объем данных в организации не может быть осмыслен отдельными сотрудниками, возникает потребность в представлении этих данных на разных уровнях абстракции так, чтобы их можно было понять и чтобы руководство организации могло принимать относительно них соответствующие решения¹⁰³.

Как стратегия данных не существует в вакууме без стратегических документов организации в целом (см. раздел 4.3), так и архитектура данных входит в целостную архитектуру предприятия (Enterprise Architecture¹⁰⁴). Она описывается с помощью комплекса проектных документов разной степени абстракции, включая стандарты, политики и прочие документы, определяющие порядок сбора, хранения, упорядочения, использования и удаления данных.

Наиболее детализированным архитектурным документом в области данных является 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 документации входят физические модели данных, но только в качестве продуктов области моделирования и проектирования, а не области архитектуры данных¹⁰⁵.

Далее наступает черед внедрения процессов и функций, **обеспечивающих доступность данных для использования и их обслуживание**. Наконец, данные используют в аналитике, за счет чего их ценность повышается (см. главу 9). Для успешной поддержки производства и использования данных, а также повышения уровня исполнительской дисциплины, организации должны осуществлять руководство данными (Data Governance; см. об этом раздел 4.4). А все названные функции в совокупности составляют систему управления данными, описанную ниже.

4.2 СОЗДАНИЕ СИСТЕМЫ УПРАВЛЕНИЯ ДАННЫМИ В ОРГАНИЗАЦИИ

Полноценное управление на основе данных невозможно без **системы управления данными**¹⁰⁶, включающей в себя отлаженные процессы работы с данными, сформированную архитектуру данных (включая

¹⁰³ DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

¹⁰⁴ Потеев П., Рудь В. Что такое архитектура организации // Стратегия цифровой трансформации: написать, чтобы выполнить. М.: РАНХиГС, 2021. URL: <https://strategy.cdto.ranepa.ru/5-1-chto-takoe-arhitektura-organizacii>

¹⁰⁵ DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

¹⁰⁶ Часто говорят также о реализации организационной функции управления данными.

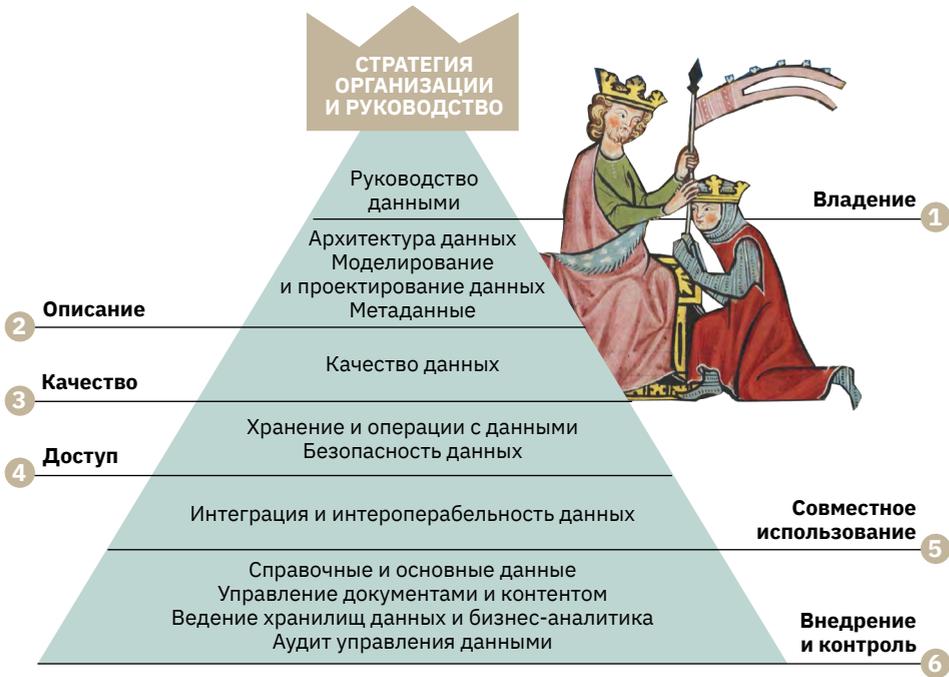


Рисунок 21. Иерархия принципов создания системы управления данными

модели данных), процессы и инструменты обеспечения качества данных, организацию хранения данных и меры по обеспечению их безопасности, а также ряд других составляющих, в основном входящих в DAMA-фреймворк и описанных в главе 1. Общий взгляд на систему управления данными представлен на рис. 21. На нем справа и слева от пирамиды расположены основные принципы создания системы управления данными в порядке приоритетности¹⁰⁷. Реализация каждого предыдущего принципа обеспечивает основу для реализации последующих.

1. Принцип владения (принадлежности). Все данные, используемые в организации, должны иметь назначенного ответственного за данные, который обеспечивает надлежащее распоряжение ими.

2. Принцип описания. Все данные должны быть соответствующим образом описаны, чтобы правильно понимать их особенности, содержание и назначение внутри организации.

3. Принцип обеспечения качества. Все данные должны быть надлежащего качества, соответствующего целям их использования в организации.

4. Принцип обеспечения доступа. Все данные по умолчанию должны быть доступны тем, у кого есть законные основания их использовать. Кроме того, данные должны быть надежно защищены от потери, повреждения или неправильного использования.

¹⁰⁷ 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, 2–16, 2016.

5. Принцип совместного использования. Все данные должны быть доступны для обмена с любой организацией или физическим лицом, имеющими соответствующие права, и совместно использоваться надлежащим образом.

6. Принцип реализации (внедрения и контроля). Должно быть реализовано управление всеми категориями данных с учетом их специфики.

Эта пирамида показывает ключевые аспекты работы с данными, о которых идет речь в навигаторе. Главы с 4 по 8 описывают (с разной степенью подробности) области, представленные в пирамиде на рис. 21, предлагая достаточно полную картину **системы управления данными** и мероприятий, входящих в **программу управления данными**.

На этапе разработки и внедрения программы управления данными необходимо направить основные усилия организации на начальные мероприятия в верхних функциональных областях (в первую очередь в той, которая соответствует принципу владения). Это мероприятия, имеющие отношение к стратегии управления данными, организационной структуре, организационным механизмам, процессам, инструментам, стандартам.

По мере внедрения постоянно действующей программы управления данными и перехода к выполнению входящих в нее проектов усилия перераспределяются с верхних областей на нижние (прежде всего на область внедрения и контроля для конкретных категорий данных). Соответственно, чем лучше будут реализованы элементы верхних функциональных областей на этапе разработки и внедрения программы управления данными (см. подробно об этом приложение А), тем меньше дополнительных усилий они потребуют впоследствии при выполнении регулярных проектов. Это позволит полностью сосредоточиться на цели этих проектов — получении максимальной выгоды от работы с данными.



Указанные принципы, в частности, легли в основу стандартов по управлению данными правительства Абу-Даби (ОАЭ)¹⁰⁸. Это один из первых всеобъемлющих (по охвату областей знаний и этапов жизненного цикла) документов в области стандартизации управления данными для правительственных организаций¹⁰⁹. Он используется в масштабной программе, охватывающей более 50 государственных ведомств¹¹⁰. Комплекс стандартов соответствует фреймворку DAMA и обеспечивает поддержку действующей в правительстве Абу-Даби политики управления данными¹¹¹.

¹⁰⁸ 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>

¹⁰⁹ Edge J. Abu Dhabi Government Publishes World's First Comprehensive Data Management Standards. 2016. URL: <https://medium.com/@berrylands/abu-dhabi-government-publishes-worlds-first-comprehensive-data-management-standards-3f0e75f0a94f>

¹¹⁰ Abu Dhabi Government Data Management Program. Abu Dhabi Systems & Information Centre (ADSIC), 2016. URL: <https://addata.gov.ae/policies>

¹¹¹ Схожий комплекс стандартов, ориентированный на фреймворк DAMA, разработан Национальным офисом управления данными Саудовской Аравии для правительственных организаций этой страны. URL: <https://www.my.gov.sa/wps/portal/snp/content/dataprotection>

4.3 СТРАТЕГИЧЕСКОЕ ПЛАНИРОВАНИЕ В ОБЛАСТИ ДАННЫХ

Данные как важнейший актив организации играют ключевую роль в ее стратегии. Поэтому важно согласовать работу с данными со стратегическими целями деятельности. Для этого организация разрабатывает **дата-стратегию**¹¹² — план использования данных для получения конкурентных преимуществ и реализации целей организации. Эта стратегия должна исходить из понимания информационных потребностей организации: 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.

Обычно дата-стратегия — это высокоуровневый документ, фиксирующий основные цели и принципы работы с данными, но она может быть и частью стратегии цифровой трансформации организации¹¹³. Ее дополняют более детальной **стратегией управления данными**, которая представляет собой программу обслуживания данных и повышения их качества, обеспечения их целостности, регулирования доступа к ним, их защиты и минимизации рисков. В ней также должны быть предусмотрены способы решения известных проблем в области управления данными¹¹⁴.



Стратегия управления данными включает в себя:

- › убедительно изложенное видение управления данными;
- › краткое экономическое обоснование с примерами;
- › руководящие принципы, ценности и перспективы управления;
- › миссию и долгосрочные цели по основным направлениям управления данными;
- › критерии успешности управления данными;
- › краткосрочные (12–24 месяца) задачи управления данными¹¹⁵;
- › описания ролей и организационных систем управления данными, включая распределение обязанностей и прав принятия решений, а также описание бизнес-процессов взаимодействия между подразделениями;
- › описания компонентов и инициатив программы управления данными;
- › приоритетную программу работ с объемами и сроками выполнения;
- › первоначальный вариант дорожной карты с разбивкой по проектам и мероприятиям.

¹¹² Дата-стратегия — калька с англ. Data Strategy. Этот термин передают по-русски и другими способами (data-стратегия, стратегия работы с данными, стратегия данных, стратегия в области данных) или используют без перевода. Чтобы сохранить различие разных видов стратегий и связь с оригинальным термином, а также учитывая распространенность в профессиональном жаргоне терминов с компонентом «дата» (дата-центр, датасет, дата-аналитик, «биг дата»), мы выбрали вариант «дата-стратегия».

¹¹³ Подробнее об этом см.: Стратегия цифровой трансформации: написать, чтобы выполнить / под ред. Е. Г. Потаповой, П. М. Потеева, М. С. Шклярук. М.: РАНХиГС, 2021. URL: <https://strategy.cdto.ranepa.ru>

¹¹⁴ Термины Data Strategy и Data Management Strategy не всегда используются строго; разные авторы могут вкладывать в них разное содержание или употреблять как синонимы.

¹¹⁵ Эти задачи должны быть конкретными, измеримыми, практически значимыми, реалистичными и привязанными к срокам (Specific, Measurable, Actionable, Realistic, Time-Bound в соответствии с принципом SMART).

Стратегия управления данными должна охватывать все области знаний, входящие во фреймворк DAMA (см. раздел 1.3) и имеющие отношение к организации. Иногда выделяют отдельные стратегии по областям управления данными, они же области знаний: стратегию руководства данными, стратегию в области качества данных, стратегию в области метаданных, стратегию в области бизнес-аналитики и т. п. Однако, если рассматривать единую стратегию управления данными как сумму всех перечисленных стратегий, можно избежать чрезмерных затрат на планирование и координацию отдельных программ, особенно в условиях недостаточной зрелости соответствующих областей управления данными¹¹⁶.

Поскольку основное назначение дата-стратегии и стратегии управления данными — поддержка стратегии организации, сроки достижения целей в области данных должны быть увязаны со сроками достижения стратегических целей бизнеса. Разработанные документы должны быть защищены, согласованы и утверждены на уровне высшего руководства организации.

4.4 РУКОВОДИТЬ, УПРАВЛЯТЬ, РАСПОРЯЖАТЬСЯ: В ЧЕМ РАЗНИЦА

Важную роль в обеспечении соответствия между дата-стратегией и стратегией организации играет руководство данными.

РУКОВОДСТВО ДАННЫМИ (Data Governance, DG) — руководство и контроль (планирование, мониторинг и обеспечение реализации) в отношении управления информационными активами организации. Руководство данными является одной из функций **управления данными** и выступает в качестве руководящей по отношению ко всем остальным его функциям.

В отличие от управления данными как практической деятельности, где главным драйвером служит извлечение ценности из информационных активов, руководство данными сосредоточено на том, как принимаются решения по данным и как должны функционировать люди и процессы, имеющие к ним отношение. Руководство данными обеспечивает надлежащее управление данными, которое, в свою очередь, служит средством достижения целей организации.

УПРАВЛЕНИЕ ДАННЫМИ (Data Management, DM) — это, согласно DAMA-DMBOK, 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 (см. также главу 1).

¹¹⁶ Aiken P., Harbour T. Data Strategy and the Enterprise Data Executive: Ensuring that Business and IT are in Synchrony in the Post-Big Data Era. Technics Publications, 2017.

Джон Лэдли в книге «Руководство данными»¹¹⁷ описывает разницу между терминами так: руководить данными нужно для того, чтобы «делать правильные вещи» (doing the right things), а управлять данными нужно, чтобы «делать вещи правильно» (doing things right).

Взаимоотношения между стратегией организации, дата-стратегией и руководством данными¹¹⁸ отражены на рис. 22. Стратегия организации задает основные направления и контекст для дата-стратегии. Руководство данными поддерживает реализацию общей стратегии организации, влияя на связанные с управлением данными аспекты ИТ-проектов. ИТ-проекты, в свою очередь, обеспечивают предоставление необходимой информации сотрудникам и партнерам организации.

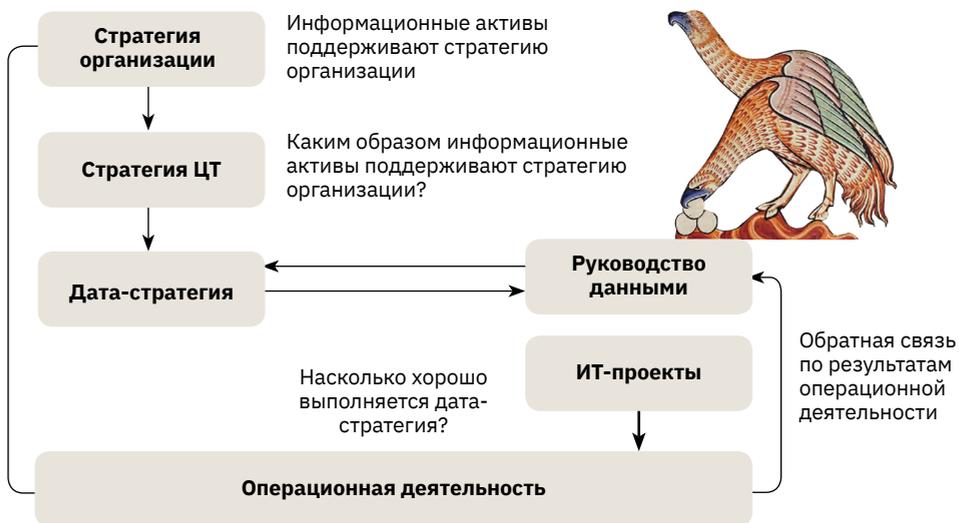


Рисунок 22. Руководство данными и дата-стратегия

Руководство данными — не разовое мероприятие. Для него требуется постоянно действующая в организации программа¹¹⁹. Наравне с программой управления данными (см. раздел 4.2) **программа руководства данными** должна предусматривать разработку политик и процедур, развитие практики распоряжения данными на разных уровнях внутри организации, инициацию мероприятий по управлению изменениями, в ходе которых

¹¹⁷ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

¹¹⁸ 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.

¹¹⁹ Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program: 2nd Edition. Academic Press, 2020.

сотрудникам будут разъяснены выгоды от внедрения руководства данными. О разработке и использовании методик и процедур, связанных с качеством и безопасностью данных, подробнее см. в разделах 7.3 и 8.3.

РАСПОРЯЖЕНИЕ ДАННЫМИ — деятельность, предполагающая ответственность за данные и процессы, обеспечивающие эффективное использование информационных ресурсов и контроль над ними¹²⁰.

Чем занимаются распорядители данных (дата-стюарды)?¹²¹

- › Часто отвечают за бизнес-гlossарий организации (см. раздел 7.2), в котором фиксируются термины, относящиеся к данным.
- › Документируют стандарты, правила качества данных, помогают убедиться в том, что в организации есть единое понимание этих правил и что они адекватно применяются.
- › Описывают данные, получаемые от разных функциональных направлений.
- › Могут найти проблемные места в управлении данными и помочь исправить ситуацию.
- › Отвечают за соблюдение политик и поддержку инициатив в области руководства данными во всех реализуемых проектах.

4.5 МОДЕЛЬ ЗРЕЛОСТИ СИСТЕМЫ УПРАВЛЕНИЯ ДАННЫМИ

Прежде чем оценивать зрелость управления данными, организация должна оценить текущее состояние (базовый уровень) своих возможностей, ресурсов, целей и приоритетов. Затем предстоит эффективно отреагировать на результаты такой оценки, определив цели, утвердив дорожную карту и наладив мониторинг прогресса.

Существующие модели оценки зрелости управления данными (Data Management Maturity Assessment, DMMA) предполагают продвижение по уровням зрелости. С переходом на очередной уровень система управления данными становится более стабильной, предсказуемой и надежной. При этом ни один уровень не может быть пропущен. Как правило, рассматриваются шесть уровней¹²² (рис. 23).

- › **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.

¹²⁰ В российской практике упор в функциях дата-стюарда делается на описании данных, а не на распоряжении ими, см. раздел 5.1 и приложение Б2.

¹²¹ См., в частности: Plotkin D. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance: 2nd Edition. Academic Press, 2020.

¹²² DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

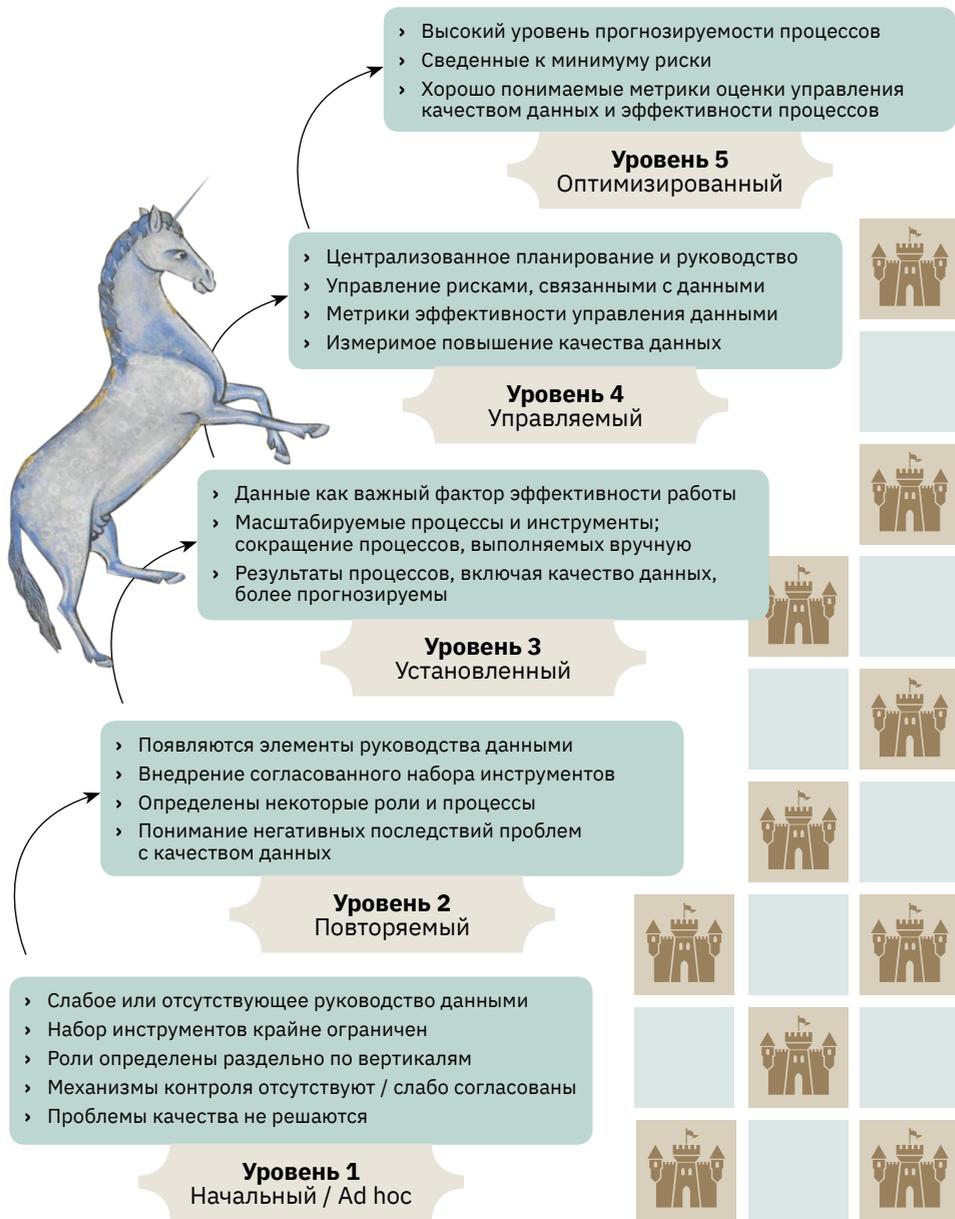


Рисунок 23. Пример модели оценки зрелости управления данными

- › **Уровень 3.** Установленный: введены и используются стандарты.
- › **Уровень 4.** Управляемый: измеряются характеристики процессов, связанных с данными, и осуществляется их контроль.
- › **Уровень 5.** Оптимизированный: обеспечена возможность измерять степень достижения целей.



«Для каждого из уровней характерна определенная корпоративная культура. Так, на уровне 1 наблюдается „феодализм данных“: каждый из элементов организационной структуры держит у себя какие-то данные и старается ими не делиться. На уровне 2 действует культура правил; появляются всевозможные ИС, в рамках которых начинается выстраивание первых справочников, но при этом наблюдается „зоопарк информационных систем“. На уровне 3 возникают инструменты, с помощью которых различные ИС или подразделения обмениваются данными с заранее оговоренными параметрами. И только на уровне 5 возникает потребность в открытых данных, а организации начинают мыслить в долгосрочной перспективе».

**Иван Добровольский,
руководитель программ ПАО «Газпром нефть»**

Существуют разнообразные методики оценки зрелости управления данными¹²³. Любая из них предусматривает наличие фреймворка из отдельно оцениваемых областей процессов управления данными. Фокус и содержание методик зависят от того, делается ли упор на общие вопросы или отраслевую специфику. Предпочтительны те методики, в которых можно отобразить используемую модель процессов на области знаний по управлению данными.



Наиболее известны, проработаны и подробно описаны (а также представлены в форме обучающих материалов и тренингов) две модели зрелости применительно к данным.

1. Модель зрелости управления данными (Data Management Maturity, DMM) Института моделирования зрелости способностей (Capability Maturity Model Institute, CMMI)¹²⁴.

2. Модель оценки способностей по управлению данными (Data Management Capability Assessment Model, DCAM) Совета по управлению корпоративными данными (Enterprise Data Management Council, EDM Council)¹²⁵.

Вторая модель ориентирована прежде всего на финансовые организации; хотя ее вполне могут применять и в других отраслях, первая модель полнее и универсальнее. В качестве приоритетной процессной области обе модели выделяют дата-стратегию и стратегию управления данными.

¹²³ В методологии оценки уровня цифровой зрелости организации, разработанной в ЦПур, данные — один из семи блоков, на основе которых производится оценка. Подробнее см. презентацию ЦПур «Цифровая зрелость». URL: <https://cpur.ru/wp-content/uploads/2020/10/Metodologiya-oczenki-czifrovoj-zrelosti-organizacii.pdf>

¹²⁴ Data Management Maturity (DMM)SM. Build, improve, and measure your Enterprise Data Management Program. // CMMI Institute. // URL: <https://stage.cmmiinstitute.com/products/dmm>

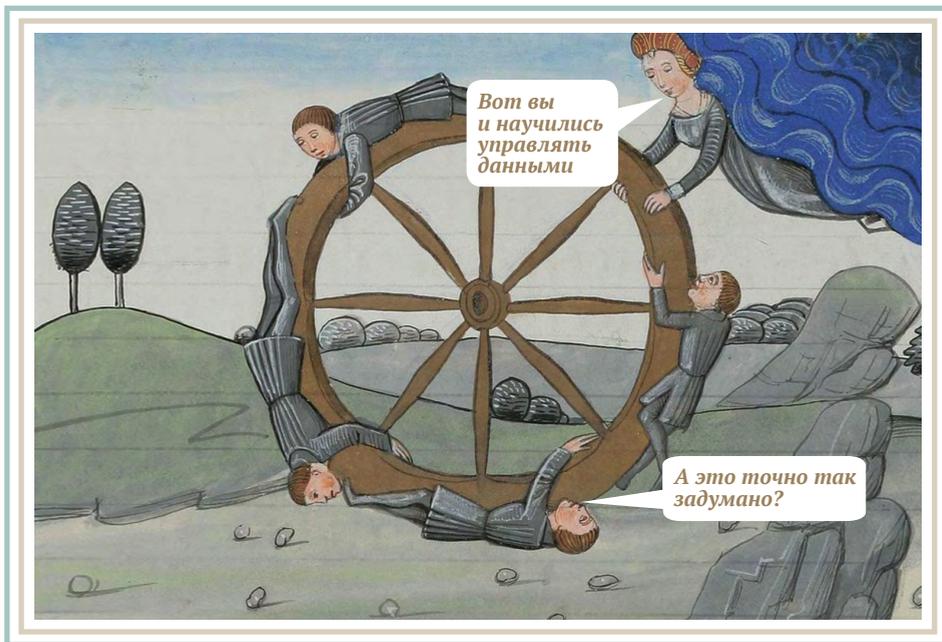
¹²⁵ DCAM Overview. EDM Council, 2021. URL: https://cdn.ymaws.com/edmcouncil.org/resource/collection/AC65DC50-5687-4942-9B53-3398C887A578/DCAM_Overview_2021_update.pdf



Рисунок 24. Пример диаграммы оценки зрелости управления данными

На рис. 24 показан пример возможной визуализации для результатов экспертной оценки уровня зрелости управления данными в организации по методологии DMM. Внешний контур задает необходимые для обеспечения конкурентоспособности организации уровни зрелости по всем областям (руководство данными, архитектура и т. д.), а внутренний отражает фактическое положение дел по этим направлениям, выявленное в ходе экспертизы. Каждая область оценивается по шкале от 1 до 5. Области с наибольшим разрывом между желаемым и текущим состояниями являются источником наибольших рисков для организации. Такая диаграмма наглядно показывает, какие области развития приоритетны для организации.

Если перед организацией стоит задача оценить зрелость управления данными, рекомендуем для начала обратиться к фреймворку (приложение А), разработанному авторами навигатора. В нем представлен краткий список вопросов для быстрой самодиагностики; ответы на них позволят понять, на каком уровне управления данными находится организация сейчас, какие у нее есть слабые и сильные стороны, куда двигаться дальше.



5. ОРГАНИЗАЦИЯ УПРАВЛЕНИЯ ДАННЫМИ И РОЛИ В КОМАНДЕ

 *Время чтения: 20 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Выстраивание работы с данными предполагает проведение первичного аудита, назначение ответственных; изменение оргструктуры и ролевой модели организации, постановку целей, выделение бюджетов.
- › При правильной организации такой проект включает как обучение сотрудников, так и оценку их цифровой грамотности.
- › Управление данными – это не управление ИТ-архитектурой, хотя данные и существуют в ИТ-системах.
- › В сфере данных есть три ключевых роли: РЦТ, руководитель по работе с данными (CDO), директор по информационным технологиям (CIO).
- › Позиция руководителя по работе с данными часто зарождается в ИТ-службе, однако лучше выстроить два независимых подразделения – ИТ и офис CDO.
- › В подразделение по управлению данными могут входить CDO, архитектор данных, специалисты по описанию данных (дата-стюарды) и по качеству данных, аналитик данных.



Н. С. Гаркуша



О. М. Гиацинтов



О. Б. Саваровская



В. А. Сазонов

5.1 КТО ОТВЕЧАЕТ ЗА ДАННЫЕ

Организация, в которой работе с данными не уделяется должное внимание, сталкивается обычно со следующими проблемами, из-за которых дублируются процессы и замедляется принятие решений:

- › данных не хватает;
- › данные появляются слишком поздно;
- › данные некачественные;
- › данные неправильно интерпретируются.

Руководителю, который планирует выстраивать работу с данными на новом уровне, важно подойти к этой задаче как к проекту, предполагающему:

- › проведение первичного аудита (со сбором реестра данных, необходимых для первостепенных задач¹²⁶);
- › назначение ответственных за различные аспекты управления данными;
- › изменение оргструктуры;
- › обновление ролевой модели организации¹²⁷;
- › и, наконец, постановку целей, создание проектов и выделение бюджетов.

Аудит можно провести самостоятельно, если организация небольшая, или с привлечением внешних консультантов; подходы к его проведению и возможные инструменты проверки — бизнес-гlossарий и каталог данных — описаны в разделе 6.2. Результатом этого проекта может стать формирование подразделения по работе с данными и дорожной карты его развития.

Кто из руководителей должен отвечать в организации за данные? Рассмотрим три руководящих роли, о которых обычно упоминают применительно к сфере данных, и попробуем определить, какие из них и в какой степени имеют дело с управлением данными.

В зоне ответственности **руководителя цифровой трансформации (РЦТ, Chief Digital Transformation Officer, CDTO¹²⁸)** находятся процессы, связанные с цифровизацией и ЦТ. В фокус его внимания попадает обновление процессов за счет внедрения цифровых решений. В круг его обязанностей, в числе прочего, входят работа с реальными данными (для аналитики, отчетности и т. д.) и стратегический взгляд на области знаний по управлению данными.

¹²⁶ Сбор реестра — крайне важный процесс. Однако в DMBOOK подчеркивается, что речь идет именно о необходимых, а не обо всех данных. Это подтверждает и опыт авторов главы: попытка собрать все данные организации может оказаться утопичной, а сбор излишних данных будет тормозить работу.

¹²⁷ В том числе переход от модели доступа в ИС к модели доступа к наборам данных.

¹²⁸ О задачах и компетенциях CDTO, отличиях этой роли от ролей CIO и CDO см.: Руководитель по цифровой трансформации // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/12/03/digital-transformation-leader/>

Руководитель по работе с данными (Chief Data Officer, CDO¹²⁹) владеет полной информацией о том, где какие данные находятся, откуда они берутся, что означают, кто ими пользуется и каким образом они могут быть предоставлены. Основная задача CDO — выстроить процессы и внедрить программные решения таким образом, чтобы обеспечить быстрый и безопасный доступ сотрудников организации к необходимым данным.

Отличие роли **директора по информационным технологиям (Chief Information Officer, CIO)** от роли CDO состоит в том, что руководитель по работе с данными ищет способы преобразования данных в ценность для организации, а директор по информационным технологиям помогает эти способы правильно реализовать, используя свой технологический опыт.

Управление данными — это не управление ИТ-архитектурой, хотя данные и существуют в ИТ-системах (так звучит один из «парадоксов данных», см. заключение). При этом во многих организациях позиция руководителя по работе с данными зарождается и существует именно в ИТ-службе, так как исторически задача описания данных и их связей находилась в ее зоне ответственности. Но такие функции не являются первоочередной задачей ИТ-службы, поэтому логичным решением будет выстроить два параллельных и независимых сервисных подразделения — **ИТ-службу и офис CDO**, — каждое из которых решает свои задачи. Подразделение по управлению данными освободит ИТ-службу от «непрофильных» аналитических запросов, поступающих от других подразделений. Модель взаимодействия подразделения по управлению данными с основными сервисами и ИТ-службой отражена на рис. 25.

В текущих структурах органов государственной власти роль CDTO отведена заместителю руководителя ведомства, заместителю руководителя субъекта РФ или заместителю руководителя учреждения. Целесообразно выделить в подчинении CDTO отдельную структуру — службу данных, которую возглавит CDO в ранге не ниже заместителя руководителя ИТ-департамента или аналогичной структуры. В субъекте РФ она может быть создана в аппарате правительства субъекта.



Рисунок 25. Взаимодействие CDO с другими функциями организации

¹²⁹ См.: The Chief Data Officer Handbook. Sunil Soares. MC Press Online, 2014 и Chief Data Officer. Synthèse des ateliers. Informatica, 2017.

Создавать новое подразделение для работы с данными необходимо; если это невозможно, стоит рассмотреть вариант расширения полномочий подразделения указанного ранга, которое ранее занималось аналитической работой.

Формирование нового направления в организации всегда сопряжено с трудностями. Далеко не все сотрудники с радостью примут структурные и процессные изменения. Большую роль в преодолении таких трудностей играет не только наделение необходимыми полномочиями офиса CDO, но и правильно выстроенная коммуникация, для чего желательно создать **комитет или совет по работе с данными**, в который входят руководители основных направлений и ключевые сотрудники. Главная задача комитета — выступать связующим звеном между производителями и потребителями данных. Как показывает опыт, при отсутствии такой поддержки новое подразделение не проходит период адаптации и его деятельность постепенно сходит на нет. Большинство руководителей не приемлет изменения подконтрольных им процессов «извне», без их непосредственного участия; комитет по работе с данными служит площадкой для вовлечения руководителей в преобразование.

В государственных структурах роль такого комитета может исполнять постоянно функционирующая **рабочая группа** из числа сотрудников нового подразделения и руководителей (или их заместителей) функциональных подразделений. Деятельность такой группы регулируется с помощью локального нормативного акта (ЛНА), в котором должен быть предусмотрен состав группы, регламент ее работы, дорожная карта с ключевыми результатами деятельности.

Структуры подразделений по управлению данными отличаются гибкостью и зависят от потребностей организации. Состав таких подразделений варьируется¹³⁰; отметим лишь ключевые роли.

Руководитель по работе с данными (CDO)¹³¹ — лидер подразделения. Он несет ответственность за все направления деятельности подразделения, обладает полномочиями, в том числе:

- › разрабатывает стратегию работы с данными внутри организации и за ее пределами;
- › формирует и поддерживает в актуальном состоянии каталог данных организации (см. раздел 7.2);
- › следит за обнаружением и инвентаризацией данных;

¹³⁰ Говоря о команде управления данными, мы здесь, например, не останавливаемся на роли ответственного за безопасность данных: исторически безопасность является отдельным подразделением, сотрудников которого привлекают к работе с данными; работу специалистов по безопасности мы рассматриваем в главе 8. Иногда в офисе CDO также выделяют роль методолога. Авторы этой главы относят написание политик и регламентов к зоне ответственности CDO, который может либо временно использовать специалистов из отдела методологии (если они есть), либо, как чаще всего бывает на практике, поручать написание политик и регламентов стюардам данных.

¹³¹ О ключевых компетенциях и профиле роли CDO см.: Руководитель по управлению данными // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/09/17/big-data/>

- › формирует процессы управления данными, их ЖЦ (включая обеспечение и повышение качества данных, см. также раздел 4.1);
- › утверждает любые изменения в описании данных (см. раздел 7.2);
- › отвечает за организацию работы комитета (совета) по работе с данными;
- › вместе с CDTO участвует в оптимизации бизнес-процессов.

Профиль роли CDO и основные компетенции этого руководителя представлены¹³² на рис. 26. Если речь идет о госструктурах, CDO должен быть основным сотрудником организации (ведомства, учреждения и т. п.).

В зону ответственности **архитектора данных**¹³³ входят концептуальное и логическое моделирование данных, построение иерархии в глоссарии (таксономий) и других справочниках; он задает правила хранения информации и учитывает требования к защите данных (об архитектуре данных см. также раздел 4.1). Он знает, как правильно описывать данные и каким образом сделать связи между ними понятными, работает в тесном сотрудничестве со стюардом данных и согласовывает его действия. Архитектор ориентируется на соглашение о моделировании, где определены основные аспекты построения моделей данных, связей между ними, ответственности и многое другое. Архитектор данных также принимает участие в изменении ИТ-архитектуры организации.

Среди ключевых функций архитектора — согласование концептуальной модели данных с функциональными подразделениями, поэтому он должен быть сотрудником рассматриваемой государственной структуры. Если остальные функции он не может выполнять самостоятельно, ему необходим помощник с соответствующими компетенциями, который может быть сотрудником подведомственного учреждения, а не самой рассматриваемой структуры.

Специалист по описанию данных (стюард данных, дата-стюард)¹³⁴ — сотрудник, ответственный за полноту и качество описания данных (метаданных). Стюард проверяет наличие описаний, вносит новые или заменяет существующие. В зарубежной практике стюарды — еще и распорядители данных (см. раздел 4.4). От объема работы, которую необходимо выполнить специалистам по описанию данных, зависит их количество в организации и формат трудоустройства. В любом случае кто-то из специалистов по описанию данных должен быть сотрудником рассматриваемой государственной структуры для поддержки всех бюрократических процессов, которые будут сопровождать основную деятельность этих специалистов.

¹³² По материалам Центра оценки и кадрового развития специалистов в области цифровой трансформации Центра подготовки руководителей и команд цифровой трансформации ВШГУ РАНХиГС.

¹³³ О роли архитектора данных и ролях других ИТ-архитекторов см.: Все про ИТ-архитектора: направления деятельности, soft-skills, специализация, заработная плата и карьера // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/05/27/it-architect/>

¹³⁴ Подробнее о компетенциях дата-стюардов см.: Data Steward: задачи, компетенции, модели работы // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/12/14/data-steward/>

Опыт работы и требования:

- › не менее 4 лет стажа государственной гражданской службы или стажа работы по специальности, направлению подготовки;
- › высшее образование.

Специальные знания и умения:

- › профессиональная терминология в сфере Big Data и Data Scientist;
- › математическая статистика, эконометрика, математическое моделирование;
- › понимание архитектуры больших данных и знание современного стека технологий работы с данными;
- › методики оценки качества и применимости моделей больших данных;
- › владение современными и классическими подходами к машинному обучению, построению нейронных сетей.

Обязанности:

- › разрабатывает, реализует концепцию развития и работы с данными внутри организации и за ее пределами;
- › обеспечивает формирование и развитие культуры работы с данными в организации;
- › обеспечивает внедрение технологий управления данными DataQuality, Big Data;
- › координирует создание аналитической платформы, разработку и поддержку ETL-процессов;
- › организует операционный процесс управления данными: инвентаризацию, сбор, систематизацию, интеграцию, использование, хранение данных;
- › организует работы по обеспечению прозрачности и собираемости, полноты и качества данных;
- › формирует и обеспечивает реализацию политики доступа к данным;
- › координирует формирование системы принятия решений на основе данных и внедрение механизмов автоматического принятия и исполнения решений с использованием технологий работы с данными;
- › организует соблюдение требований по информационной безопасности в рамках всего жизненного цикла данных.



Рисунок 26. Профиль роли руководителя по управлению данными

Основная задача **специалиста по качеству данных** — контролировать соответствие требований, предъявляемых к данным пользователями, реальному уровню качества; выстраивать единую политику и подходы к управлению качеством данных. Если такого специалиста в организации нет, частые и бессистемные проверки качества данных порождают дублирование требований и дополнительные затраты на инфраструктуру.

В число задач специалиста по качеству данных обязательно входят:

- › создание методики оценки качества данных;
- › формирование и контроль исполнения регламента оценки качества данных;
- › ведение реестра правил проверки и обеспечения качества данных;
- › работа с показателями и критериями качества.

Функция специалиста по качеству данных может относиться как к подразделению по управлению данными, так и к ИТ-службе. В случае работы в службе ИТ такой специалист может отвечать за создание и изменение правил проверки и обеспечения качества данных, хотя специалисту по качеству целесообразнее взять на себя правильную постановку ИТ-специалистам задач по разработке правил качества.

Эксперт по данным в функциональном подразделении (Subject Matter Expert, SME) — это сотрудник подразделения, знающий свою предметную область; при этом функционально он может подчиняться CDO. Такой эксперт

- › тесно взаимодействует со специалистами по описанию данных;
- › направляет дата-стюарду и архитектору данных информацию о терминах, связях, логике формирования показателей;
- › контролирует работу по заполнению описаний и предоставляет сведения об изменениях во внутренних требованиях, регламентах, процессах.

В государственных структурах эксперты по данным входят в состав рабочей группы по управлению данными и знают, как данные сформировались в их подразделении и как они там используются для принятия решений. Важно, чтобы на такую роль были назначены сотрудники, в обязанности которых уже входит работа с данными всего подразделения и аналитика.



Объединять роли специалиста по описанию данных и эксперта по данным в функциональном подразделении не рекомендуется. Специалисты подразделений, которые обычно и выступают в роли экспертов, не хотят выполнять дополнительную работу. К тому же роль эксперта предполагает наличие особых, нишевых компетенций; кроме того, в разных подразделениях требуется разный уровень детализации описаний данных.

Потребность в роли **аналитика данных** появляется в том случае, когда в обязанности специалиста по описанию данных (стюарда) входит описание требований пользователей к ИТ-службе. Если стюард описывает их на концептуальном «языке», то аналитик переводит их на язык технический. Аналитики описывают логическую модель и отправляют в ИТ-службу и инженерам данных (см. о них раздел 6.1) информацию о разработке, а потом проверяют правильность разработанных процессов. Роль аналитика данных следует отличать от роли аналитика-исследователя¹³⁵ (см. главу 9), которая, как и роль дата-инженера¹³⁶, чаще всего не относится непосредственно к команде CDO. Аналитики-исследователи обычно работают в функциональных подразделениях, дата-инженеры, как и разработчики — в подразделении ИТ.

Некоторые сотрудники подразделения по управлению данными (кроме CDO, архитектора данных и одного-двух специалистов по описанию данных) могут быть выведены за штат. Иногда в госструктурах такая организация работы позволяет внештатным сотрудникам заниматься исключительно управлением данными, не отвлекаясь на бюрократические процедуры.



Если в организации есть офис CDO, сотрудник функционального подразделения направляет ему запрос по электронной почте, через внутреннюю систему или заказав данные в бизнес-гlossарии. Всю остальную работу делает подразделение по управлению данными. Специалисты по описанию данных принимают запрос, выясняют, описаны ли запрашиваемые данные, проверяют уровень их качества; в случае недостатка информации обращаются к эксперту по данным того подразделения, из которого пришел запрос. Специалисты по описанию данных будут отвечать за поиск и описание информации, аналитики — за постановку задач для ИТ и контроль их реализации, эксперты — за обеспечение специалистов по описанию информацией по своему функциональному направлению. Примером успешной реализации такого подхода в своих организациях могут служить ПАО «Газпром нефть», банк ВТБ и Центральный банк Российской Федерации.

На первый взгляд может показаться, что появление новой функции усложнит взаимодействие внутри организации. Но на самом деле с появлением подразделения по управлению данными функциональные подразделения не только начинают получать качественную оперативную аналитику, но и экономят 40–60% времени¹³⁷.

¹³⁵ Подробнее о роли аналитика-исследователя и ее отличиях от роли исследователя данных (Data Scientist) см.: Баттл востребованных профессий: из Data Analyst в Data Scientist или наоборот // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/02/25/battl-vostrebovannyh-professij-iz-data-analyst-v-data-scientist-ili-naoborot/>

¹³⁶ О ключевых требованиях к навыкам в описаниях вакансий дата-инженеров и о плане развития компетенций дата-инженеров см. Data Engineer: анализ вакансий 2021 года и план развития // Центр оценки и кадрового развития специалистов в области цифровой трансформации. URL: <https://cdto.work/2021/01/28/data-engineer-analiz-vakansij-2021-goda-i-plan-razvitija/>

¹³⁷ Brocchi Ch., Grande D., Rowshankish K., Saleh T., Weinberg A. Designing a data transformation that delivers value right from the start. McKinsey & Company, 2018. URL: <https://www.mckinsey.com/industries/financial-services/our-insights/designing-a-data-transformation-that-delivers-value-right-from-the-start>

5.2 АЛГОРИТМ СОЗДАНИЯ ЭФФЕКТИВНОЙ КОМАНДЫ

Чтобы сформировать подразделение, которое будет приносить реальную пользу, руководителю важно действовать поэтапно, не забывая о создании процессов и методик работы с данными.

1. Инициировать проект создания новой структуры в организации, проработать оптимальную структуру команды CDO.

2. Наделить новое подразделение необходимыми полномочиями, сформировав перечень полномочий и утвердив их локальным нормативным актом (ЛНА). При этом нужно понимать, что часть необходимых полномочий ранее была закреплена за другими подразделениями (как функциональными, так и ИТ). Далее последовательно создать ЛНА для управления данными и модернизировать существующие ЛНА в части, касающейся управления данными организации (извлекая оттуда полномочия и меняя оргструктуру).

3. Документировать и структурировать данные, которыми управляет организация, поскольку управление данными может быть зафиксировано в разнородных ЛНА, описано в документации на ИС или организовано в ходе исполнения разовых поручений. Для полноценной комплектации нового подразделения сотрудниками, для модернизации всех ЛНА, которые касаются управления данными, и для планирования деятельности подразделения необходимо провести аудит.

4. Чтобы организовать работу по первым трем пунктам, а также с самого начала прояснить роль нового структурного подразделения, следует создать **совет (комитет) по работе с данными**, определить его цели и периодичность встреч. Действовать нужно быстро: в госсекторе и больших коммерческих организациях очень долгий процесс принятия решений из-за длинных бюрократических цепочек, поэтому на само выполнение решения остается меньше времени, чем ожидается.

5. Управление данными — это не труд одного человека, особенно в госсекторе, где много влияющих подразделений. В первое время неминуемо возникнет проблема разграничения полномочий между подразделениями, поэтому важно проработать **структуру управляющих органов**, которые будут работать с данными.



6. ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ ОБРАБОТКИ ДАННЫХ

 *Время чтения: 15 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Дата-инжиниринг — область знаний, связанная с разработкой, тестированием и поддержкой инфраструктуры работы с данными.
- › Цель инженера данных — внедрить систему (платформу), куда автоматически загружаются данные, где они получают доступную для конечных пользователей форму и где пользователи могут самостоятельно проверять гипотезы и решать свои задачи, используя правильные метрики.
- › В архитектуре аналитической рабочей среды организации выделяются пять слоев: источники данных, обработка данных, хранение данных, наука о данных и бизнес-аналитика.
- › Слой обработки данных выделен для обозначения операций, которые идут либо в потоковом (непрерывном) режиме, либо в пакетном (с перерывами)
- › Слой хранения включает традиционное хранилище данных, оно же Data Warehouse, и хранилище больших данных (озеро данных); другой вариант — современное хранилище, объединяющее DWH и озеро данных (платформа данных).

6.1 ЗАДАЧИ ИНЖЕНЕРА ДАННЫХ

Когда в организации создана команда для управления данными (см. раздел 5.1), встает вопрос о том, как подступиться к большим объемам сырых данных: как их правильно собирать, обрабатывать и хранить.

ДАТА-ИНЖИНИРИНГ — область знаний, связанная с разработкой, тестированием и поддержкой инфраструктуры работы с данными: баз данных, хранилищ, систем массовой обработки и подготовки данных. К ней относится построение процессов извлечения, загрузки, преобразования, хранения и подготовки данных для систем-потребителей: аналитики, алгоритмов машинного обучения, науки о данных (Data Science) и т. д.

Инженеры данных — это инженеры-программисты, которые отвечают за построение конвейеров данных (Data Pipelines) для объединения информации из разных систем-источников. Они интегрируют, консолидируют и очищают данные, строят хранилища данных, извлекают и структурируют данные для использования в аналитических решениях (приложениях, системах). Объем данных, с которыми работает инженер, зависит от характеристик организации, особенно от ее размера. Некоторые отрасли обрабатывают данные более интенсивно (например, здравоохранение, розничная торговля, банковский сектор). Чем крупнее организация, тем больше данных, за которые отвечает инженер.



Отличие дата-инженера от аналитика и исследователя данных можно описать метафорой¹³⁸. Прежде чем приготовить морковный пирог, нужно собрать морковь, привезти ее с огорода, помыть и почистить, и роли в этом процессе распределяются так: инженер собирает морковь, привозит ее, моет и чистит, а аналитик (или исследователь) готовит из нее пирог.

Экосистема больших данных (см. раздел 2.3) может включать самые разные технологические компоненты¹³⁹, показанные на рис. 27, и прежде всего распределенные файловые системы. Они работают на нескольких серверах сразу, способны хранить файлы, которые не умещаются на диск отдельного компьютера, ориентированы на параллельную обработку файлов (одновременно на нескольких компьютерах) и легко масштабируются¹⁴⁰.

¹³⁸ Кто такие дата-инженеры, и как ими становятся? // Хабр. URL: <https://habr.com/ru/company/otus/blog/452670/>

¹³⁹ Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2018.

¹⁴⁰ Наиболее популярной распределенной файловой системой является Hadoop File System.

Авторы главы:



В. А. Сазонов



Н. В. Скворцов



Р. С. Стрекаловский



Рисунок 27. Экосистема больших данных

Хранение огромных объемов данных предполагает использование ПО (систем управления базами данных, СУБД), предназначенного для управления такими данными и формирования запросов к ним. Традиционные реляционные базы данных^{141, 142}, использующие язык запросов SQL (такие как Oracle, MySQL и т. п.), плохо справляются с большими объемами записи и чтения данных, хотя отлично работают с задачами поиска данных в заданной структуре. Поэтому появились новые типы СУБД на основе нереляционных технологий, объединенные в категорию NoSQL¹⁴³.

Данные в распределенной файловой системе перемещаются от источников к потребителям с помощью специальной инфраструктуры интеграции данных. Когда же данные доходят до потребителя, начинается их обработка и аналитика с целью извлечения из них полезной информации и знаний.

6.2 ОСНОВНЫЕ ОПЕРАЦИИ С ДАННЫМИ: СБОР, ХРАНЕНИЕ, ОБРАБОТКА

Цель инженера данных — внедрить систему (платформу), куда автоматически загружаются данные, где они трансформируются в доступную для конечных пользователей (или бизнес-пользователей) форму и где пользователи могут самостоятельно работать с данными, проверять гипотезы и анализировать свои задачи (бизнес-задачи), используя правильные метрики.

¹⁴¹ База данных — совместно используемый набор логически связанных данных (и описание этих данных), предназначенный для удовлетворения информационных потребностей организации. См.: Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. М.: Вильямс, 2003.

¹⁴² Реляционные СУБД поддерживают реляционную модель данных (от англ. relation — ‘отношение’, ‘зависимость’, ‘связь’). База данных, построенная в соответствии с такой моделью, представляет собой совокупность взаимосвязанных таблиц (в терминах теории реляционных баз данных — отношений), каждая из которых содержит сведения об объектах определенного типа (товарах, клиентах и т. п.).

¹⁴³ От англ. not only SQL — не только SQL.

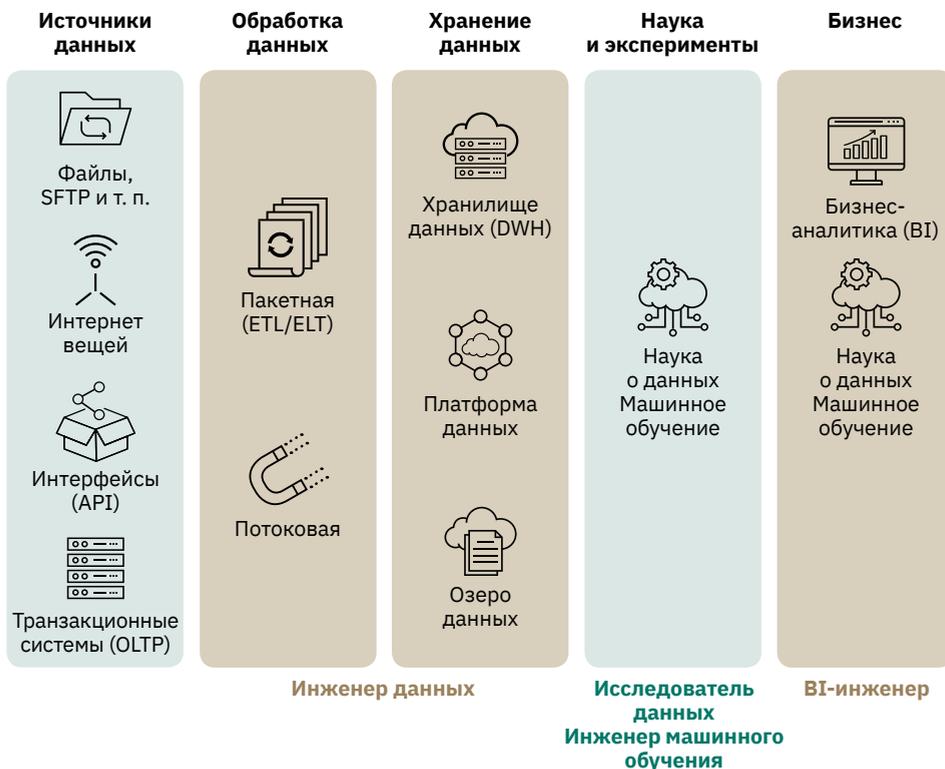


Рисунок 28. Архитектура аналитической рабочей среды организации

На рис. 28 представлена архитектура аналитической рабочей среды организации, в которой выделяются пять слоев. В слой **источников данных** могут входить различные приложения, подключаемые по API¹⁴⁴, а также датчики, внешние устройства и другие источники данных, которые подключаются напрямую или с помощью сетевых протоколов. Также данные могут поступать из систем оперативной обработки транзакций (OnLine Transaction Processing, OLTP), поддерживающих операционную деятельность организации.

Слой **обработки данных** выделен для обозначения операций, которые идут в потоковом режиме (непрерывно) или в пакетном (с перерывами). В пакетном режиме данные могут двигаться либо по схеме ETL (Extract, Transform, Load; извлечение → преобразование → загрузка), либо по схеме ELT (Extract, Load, Transform; извлечение → загрузка → преобразование).

Слой **хранения** включает традиционное хранилище данных (ХД), оно же Data Warehouse (DWH), и хранилище больших данных (озеро данных); другой вариант — современное хранилище, объединяющее DWH и озеро данных (платформа данных). Два последних слоя выделены для обозначения деятельности в областях **науки о данных** и **BI**.

¹⁴⁴ API (Application Programming Interface, интерфейс прикладного программирования) — описание способов, которыми одна компьютерная программа может взаимодействовать с другой.

Рассмотрим подробнее слою обработки и хранения. В DWH собираются данные для бизнес-аналитики (см. раздел 9.5.1); перед хранением они должны быть очищены и упорядочены. Эти данные уже обработаны, поэтому их легче использовать для высокоуровневого анализа: к ним легко применимы инструменты BI, и такими хранилищами пользуются не только аналитики и специалисты в области работы с данными. В озере данных собираются необработанные большие данные в изначальном собственном формате. Озеро может хранить данные различной структуры. Каждый сохраненный элемент помечен уникальным идентификатором и снабжен метаданными, чтобы при необходимости его можно было легко запросить. Данные в озере применяются для решения задач **науки о данных** (см. раздел 9.3) и предназначены прежде всего для исследователей данных.



В центре обработки данных «Газпром нефти» в Санкт-Петербурге создано «умное» озеро данных. Им пользуются аналитики бизнес-подразделений в 29 регионах РФ и странах СНГ, различные системы внутри компании и внешние партнеры¹⁴⁵. «Умное» озеро основано на сложной современной аналитической архитектуре¹⁴⁶. Компоненты обработки, хранения и анализа данных (озеро данных, хранилище данных, научная лаборатория данных) работают в тесной интеграции с компонентами руководства данными (каталогом данных, управлением качеством данных, бизнес-гlossарием, пользовательским порталом обзора данных). В решение включены более 140 внутренних и внешних источников данных.

В настоящее время известно несколько перспективных архитектурных концепций построения платформы данных, в частности Modern Data Architecture, Lambda Architecture и Data Mesh Architecture¹⁴⁷.

- › Modern Data Architecture соединяет преимущества DWH и озера данных. Реализация во многом зависит от главного инженера проекта.
- › Lambda Architecture позволяет решать задачи, связанные с обработкой в режиме реального времени, за миллисекунды.
- › Data Mesh Architecture объединяет пакетную и потоковую обработку данных (стриминг), а хранит данные в облаке. Организации могут анализировать данные в режиме реального времени, снизив при этом затраты на управление инфраструктурой хранилища.

При построении хранилищ и озер целесообразно следовать передовым практикам¹⁴⁸, методологиям и регламентам^{149, 150}.

¹⁴⁵ Первое в России внедрение «умного озера данных» — комплексной платформы обработки, хранения и анализа данных с интегрированным Data Governance // Global CIO. URL: <https://globalcio.ru/projects/3040/>

¹⁴⁶ Building Data Lakes Successfully — Part 1 — Architecture, Ingestion, Storage and Processing // Gartner. URL: <https://www.gartner.com/en/documents/3991474-building-data-lakes-successfully-part-1-architecture-ing>

¹⁴⁷ Пять подходов к построению современной платформы данных // Arenadata. URL: <https://arenadata.tech/about/blog/5-ways-to-dwh/>

¹⁴⁸ См. обзор актуальных технологий в области дата-инжиниринга и управления данными: Gartner® Hype Cycle™ for Data Management, 2021 // Databricks. URL: <https://databricks.com/p/ebook/gartner-hype-cycle-for-data-management-2021>

¹⁴⁹ DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

¹⁵⁰ Gorelik A. The Enterprise Big Data Lake Delivering the Promise of Big Data and Data Science. O'Reilly Media, Inc., 2019.

Рассмотрим подробнее деятельность инженера данных во время обработки и сохранения данных на примере операций, выполняемых в ходе процесса ETL.

1. Извлечение данных. На этом этапе данные извлекаются из одного или нескольких источников и подготавливаются к преобразованию. Для корректного представления данных после их загрузки в хранилище нужно извлечь из источников не только сами данные, но и информацию, описывающую их структуру, из которой будут сформированы метаданные для хранилища.

2. Преобразование данных состоит из нескольких шагов. Данные из различных источников могут отличаться соглашениями о назначении имен полей и таблиц, порядком их описания, форматами, типами и кодировкой. Перед передачей в хранилище их нужно свести к единой структуре.

«Наибольший интерес для анализа, как правило, представляют данные, обобщенные по некоторому интервалу времени, по группе клиентов, товаров и т. д. Такие обобщенные данные называются агрегированными (агрегатами), а сам процесс их вычисления — агрегированием»¹⁵¹. **Агрегирование** проводится в процессе преобразования. Чтобы снизить избыточность данных и уменьшить их объем при хранении в источниках, часто используют специальные кодировки. «Так, наименования объектов, их свойств и признаков могут храниться в сокращенном виде. В этом случае перед загрузкой данных в хранилище требуется выполнить **перевод** таких сокращенных **значений** в более полные и, соответственно, понятные»¹⁵².

В процессе загрузки в хранилище может понадобиться вычисление некоторых новых данных на основе существующих, что обычно сопровождается созданием новых полей.

Наконец, выполняется **очистка данных**. «Грязные» данные — одна из важнейших и трудно формализуемых проблем аналитических технологий. Существует несколько причин, по которым данные нуждаются в очистке; чаще всего они связаны с нарушением структуры данных, а также ошибочными значениями полей. Иногда очистка данных необходима для обеспечения безопасности данных: в процессе очистки чувствительная информация (например, ПДн) заменяется на уникальные случайные значения, оставляя возможность построения всевозможной аналитики (см. об этом также в главе 8).

3. Загрузка данных. Процесс загрузки заключается в переносе данных из промежуточных таблиц в структуры хранилища данных. От выверенности процесса загрузки данных во многом зависит время, которое требуется для полного цикла их обновления в хранилище, а также их полнота и корректность (см. раздел 7.3).

¹⁵¹ Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям: учебное пособие. 2-е изд., испр. СПб.: Питер, 2013.

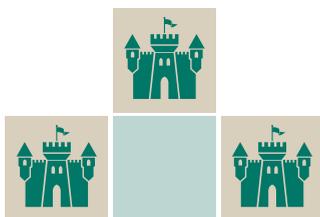
¹⁵² Там же.

Описанный здесь набор 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, может возникнуть необходимость в осуществлении **группировки** или разгруппировки данных (объединении или разъединении данных по какому-либо признаку), **нормализации** (преобразовании диапазона изменений числового признака в другой, более удобный для применения в процессе анализа), **квантовании** (разбиении диапазона возможных значений числового признака на заданное количество интервалов и присвоении попавшим в них значениям номеров интервалов или иных меток) и т. п.

6.3 ПОДГОТОВКА К ОБЕСПЕЧЕНИЮ ПРОЦЕССОВ ОБРАБОТКИ ДАННЫХ В ОРГАНИЗАЦИИ: ЧЕК-ЛИСТ

С помощью этого краткого списка вы сможете убедиться, что не пропустили ни одного из ключевых этапов, предшествующих созданию работающей архитектуры хранения и обработки данных. Какие действия важно выполнить?

- 1. Определить источники данных** и понять, на каком этапе жизненного цикла (см. раздел 4.1) эти данные находятся.
- 2. Структурировать данные из разных источников.** Создайте каталог данных (см. раздел 7.2) — один из ключевых инструментов управления данными (наряду с бизнес-глоссарием). Если источников много, опишите их, выбрав удобную нотацию.
- 3. Определить, какой объем данных и с какой скоростью необходимо собирать и обрабатывать,** — это позволит спроектировать нужную архитектуру хранения и обработки.
- 4. Обсудить с РЦТ и со всеми руководителями, какие именно данные для них приоритетны.** Если уже есть четкие, понятные структуры данных в исходниках и их нужно собирать и агрегировать для аналитики и отчетности, то, возможно, достаточно разворачивания хранилища данных. На рынке есть как системы, которые даже на локальной станции могут агрегировать все нужные источники от Excel-таблиц до корпоративных информационных систем (таких как, например, «1С:Предприятие»), так и программы, которые позволяют собирать неструктурированные данные.
- 5. Описать дальнейшие этапы жизненного цикла данных.** Возможно, после обработки данные будут опять возвращаться в системы, но уже как более проверенные, валидированные. Возможно, они будут использоваться только для аналитики или построения моделей для ИИ.



БЛОК III
ВЫСТРАИВАЕМ
ПРОЦЕССЫ



7. ПРОБЛЕМА КАЧЕСТВА ДАННЫХ

 *Время чтения: 15 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Различают несколько уровней зрелости качества данных. На первом — данные отсутствуют, на втором — данные есть, но они некачественные, на последнем — правильно интерпретируемые данные служат основой для принятия управленческих решений.
- › Работа над качеством данных начинается с описания их текущего состояния в организации. Важно выяснить, как устроены процессы управления данными, кто владеет этими процессами и участвует в них, есть ли регламенты, политики и программные средства.
- › Результаты аудита фиксируются в бизнес-глоссарии и каталоге данных.
- › Как правило, выделяют следующие метрики качества данных: полнота, своевременность и волатильность, точность, валидность, согласованность, наличие и доступность.
- › Данные должны работать. Чем чаще данные используют, тем больше вероятность выявить ошибку.
- › Чтобы качество реально повышалось, а данные становились активом, необходимо встроить жизненный цикл данных во все ключевые процессы организации.

7.1 УРОВНИ ЗРЕЛОСТИ КАЧЕСТВА ДАННЫХ В ОРГАНИЗАЦИИ

Работа над качеством данных начинается с описания их текущего состояния в организации. Нет смысла строить аналитические модели и системы поддержки принятия решений, не имея собственно данных. Также нет смысла обучать нейросети на некачественных данных (если только они не являются частью обучения): получится некачественная модель.

Данные и их качество не существуют сами по себе, они связаны с процессами деятельности организации. Описывая текущий статус работы с данными в организации, мы выясняем, какими объектами управляем и какие процессы над ними совершаются, какие требования предъявляются к качеству данных и в каких ресурсах эти данные аккумулируются, какая проверка качества данных (при их наличии) ведется по базовым метрикам качества и, наконец, какая аналитика готовится на основе данных и как она используется при принятии решений.



«Разработчики практических решений, основанных на технологиях ИИ, и научные сотрудники сталкиваются с проблемами качества данных, ограничениями использования. Решать проблему качества государственных данных (как открытых, так и тех, которые могут быть полезны даже при ограничении к ним доступа) предстоит для всех пользователей. Чтобы достичь успеха, необходимо ориентироваться на пользователей, воспринимая их как клиентов, заказчиков качества данных»¹⁵³.

**Мария Шклярчук, академический директор
Центра подготовки РКЦТ**

Зрелость качества данных можно рассматривать в контексте оценки зрелости управления данными в целом (см. раздел 4.5). Можно выделить пять уровней зрелости качества данных.

1. Данных нет. Необходимо, чтобы они появились

Под отсутствием данных мы понимаем как их фактическое отсутствие, так и наличие исключительно в бумажном виде или, например, в формате электронных изображений, при котором

Авторы главы:



Я. Э. Гейн



О. М. Гиацинтов



А. А. Малахов



О. Б. Саваровская

¹⁵³ Из предисловия к аналитическому обзору ЦПУР «Модели институционализации доступа исследователей к данным государства». URL: https://cpur.ru/research_pdf/PP_data_access_models_2021_FIN_ped.pdf

автоматизированная обработка информации без предварительной трудоемкой подготовки невозможна. Об отсутствии данных можно говорить и в ситуации, когда имеющиеся данные неполны, то есть не описывают всю полноту объектов управления одного типа.

Весьма распространен «феодализм данных»: данные вроде бы есть в одних подразделениях, но не попадают в другие в полном объеме, несмотря на (казалось бы) сквозные процессы в организации.

Главная цель на этом этапе — наличие данных. Создавать ИС ведения информационного ресурса с версионностью, проверками и валидацией данных в таких случаях не всегда оправдано — это долго и дорого, предпочтительнее сделать два технически простых, но очень важных и организационно сложных шага:

- › сформировать электронный информационный ресурс, где будут содержаться данные об объектах управления. Важно корректно описать состав полей — модель данных, максимально используя при этом справочники и классификаторы, минимально — ручной ввод. При этом вовсе не обязательно создавать сложную и дорогую информационную систему: корректно сформированный табличный файл (Excel, Open document table) способен закрыть базовую потребность в наличии данных;
- › изменить процессы в организации так, чтобы любое действие, меняющее информацию об объектах управления, фиксировалось в ранее созданном информационном ресурсе.

2. Данные есть, но они некачественные. **Осознанная работа над качеством данных**

На этом этапе начинают работать базовые метрики качества данных (см. раздел 7.3). По сути, каждая из метрик указывает на возможные проблемы. Разные проблемы требуют разных подходов, кратко остановимся на двух основных.

- › Проблемы качества отдельных записей: унаследованные ошибки в данных (в отдельных случаях — человеческий фактор), в том числе ошибки работы каких-то функций, случившиеся нетиповые ситуации и т. п. Необходимо выстроить процессы работы по таким инцидентам, а для государственных данных — регламентные процедуры отработки этих ошибок.
- › Системные проблемы при организации работы с данными: массовые ошибки, которые типизируются и вызваны проблемами в процессах. Для их устранения необходима дальнейшая системная работа по изменению процессов, встраивание в них жизненного цикла данных, выделение ответственных сотрудников. Системные проблемы могут быть связаны также с проблемой справочников и классификаторов (см. раздел 3.2.1), некорректными моделями данных.



К качеству данных, участвующих в межведомственном обмене (см. раздел 3.1), предъявляются повышенные требования: получатели таких данных вправе не расходовать время и ресурсы на их дополнительную проверку и очистку. Рекомендации НСУД¹⁵⁴ способствуют раннему выявлению и недопущению таких распространенных ошибок в данных, как, например:

- › хранение числовых данных и дат в текстовых полях (ошибки типа данных);
- › хранение в одном поле всех реквизитов адреса объекта (нарушение атомарности данных);
- › несоответствие текста требованиям Unicode и непредставленность его в кодировке UTF-8 (ошибка заполнения тестовых полей).

Особое внимание в рекомендациях уделено использованию справочников и классификаторов, а также проблеме дублирования данных.

3. Есть качественные данные. Обогащение данных из собственных и внешних источников

Организация может уже осознавать потребность в данных для поддержки принятия управленческих решений, но не иметь полных данных. Создать новое знание, которое невозможно получить из каждого набора данных в отдельности, способно обогащение данных (см. разделы 8.1 и 8.2). На этом этапе нужно уделить внимание двум ключевым аспектам:

- › гармонизации данных — использованию единых, унифицированных справочников и классификаторов, данных из первоисточников с единообразным описанием данных (о семантической интероперабельности см. также раздел 3.2.1) и едиными форматами представления данных (технической интероперабельностью);
- › расширенным проверкам качества данных — дальнейшему повышению качества данных, в том числе за счет кросс-проверок, поскольку при обогащении данных возможно накопление ошибок (см. раздел 8.2).

4–5. Есть качественные, обогащенные данные. Запрос на аналитику для принятия управленческих решений

Управление качеством данных на этих этапах — уже хорошо организованный процесс, который встроен в основные процессы организации. Качество необходимо поддерживать, отслеживать актуальность правил контроля качества данных, следить за целостностью справочников и классификаторов, чтобы получать корректную аналитику (см. главу 9).

Финальный пятый этап предполагает, что правильно интерпретируемые данные становятся основой для принятия управленческих решений.

¹⁵⁴ Методические рекомендации по обеспечению качества данных, предназначенных для межведомственного обмена. URL: <https://nsud.info.gov.ru/ifp/portals/documents/49>

7.2 АУДИТ УПРАВЛЕНИЯ ДАННЫМИ. БИЗНЕС-ГЛОССАРИЙ И КАТАЛОГ

Рассмотрим подробнее формат анализа текущего состояния управления данными в организации. В ходе проверки важно выяснить, как устроены процессы управления данными, кто ими владеет и участвует в них, есть ли в этой области какие-либо регламенты, политики и программные средства. Для фиксации результатов проведенного анализа можно воспользоваться двумя инструментами (точнее, типами инструментов): бизнес-гlossарием и каталогом данных.

БИЗНЕС-ГЛОССАРИЙ — информационная система для описания данных: терминологии, концептуальной и логической моделей, реестра правил качества данных, политик, регламентов, процессов и связей между ними. Информация вносится в гlossарий независимо от того, с чего будет начат аудит, главное — избегать ее дублирования.

Среди функций бизнес-гlossария должны присутствовать:

- › описание терминологии, которая применяется в отчетах, выгрузках и других типах работы с данными;
- › описание процессов и проектов работы с данными;
- › ведение логической модели данных (графическое представление логической структуры) и ее связей с терминологией;
- › учет связанных с данными распоряжений и политик (см. табл. 3), правил качества данных;
- › описание данных в контексте систем, межсистемных интерфейсов, наборов данных и связей между ними;
- › перечень сотрудников, отвечающих за данные;

Таблица 3. Пример учета регламентов в бизнес-гlossарии с указанием ответственных лиц

Имя объекта	Роль	Имя	Организационная единица
Отчет о прибыли компании по продажам контрагентов и клиентов за период	Владелец системы	А. А. Коновалов	Блок ИТ
Отчет о прибыли компании по продажам контрагентов и клиентов за период	Ключевой пользователь	В. Н. Сердюкин	Офис CDO
Отчет о прибыли компании по продажам контрагентов и клиентов за период	Ключевой пользователь	Д. Ю. Сидоров	Офис CDO
Отчет о прибыли компании по продажам контрагентов и клиентов за период	Ключевой пользователь	И. И. Иванов	Блок финансов

- › ведение ЖЦ данных;
- › контроль над изменениями хранимой информации;
- › формирование заказа на поставку данных (функция «супермаркета данных»).

В ходе аудита отдельного внимания требуют программные разработки, которые так или иначе связаны с работой с данными (аналитика, отчеты, выгрузки). Для этого предназначен каталог данных.

КАТАЛОГ ДАННЫХ — ИС выявления метаданных, используемых для хранения и перемещения данных и построения взаимосвязей между ними. Основные функции каталога данных:

- › извлечение физических метаданных систем хранения, интеграции, трансформации, передачи, обеспечения качества данных;
- › обнаружение доменов (допустимых значений данных) с помощью правил или ИИ для привязки к терминологии глоссария;
- › автоматизированная привязка терминологии бизнес-глоссария к объектам физических метаданных;
- › поиск данных.

В процессе аудита в каталог поступают сведения из систем хранения или аналитики данных (см. главу 9), а также от тех процессов обработки данных, которые описаны в разделе 6.2. Таким образом возникает визуализация взаимосвязей (см. рис. 29) между данными, существующими в организации, — схема или путь данных (Data Lineage; см. также раздел 4.1). Каталог — инструмент отслеживания соответствий между регламентами/требованиями и программной реализацией.

Аудит управления данными в организации — небыстрый процесс. Рекомендуется разделить его на несколько небольших этапов и последовательно описывать ситуацию в каждом функциональном блоке, отсылая информацию о проблемах руководителям ЦТ. Ведение информации в глоссарии и использование каталога данных применимо только для аудируемых функциональных блоков и подразделений. После того как будет собрана информация по всем блокам и подразделениям, появится возможность:

- › сформулировать задачи и цели ЦТ в организации;
- › оценить состав команды и сформировать ее;
- › понять, какие нужны регламенты и политики, какие решения нужно принять руководству, чтобы работа с данными велась эффективно;
- › разработать план действий.

Аудит выявит, кто является владельцем инвентаризированных данных, кто — потребителем, а кто может помочь в работе по поддержке и улучшению качества данных; он также позволит создать реестр данных, которые

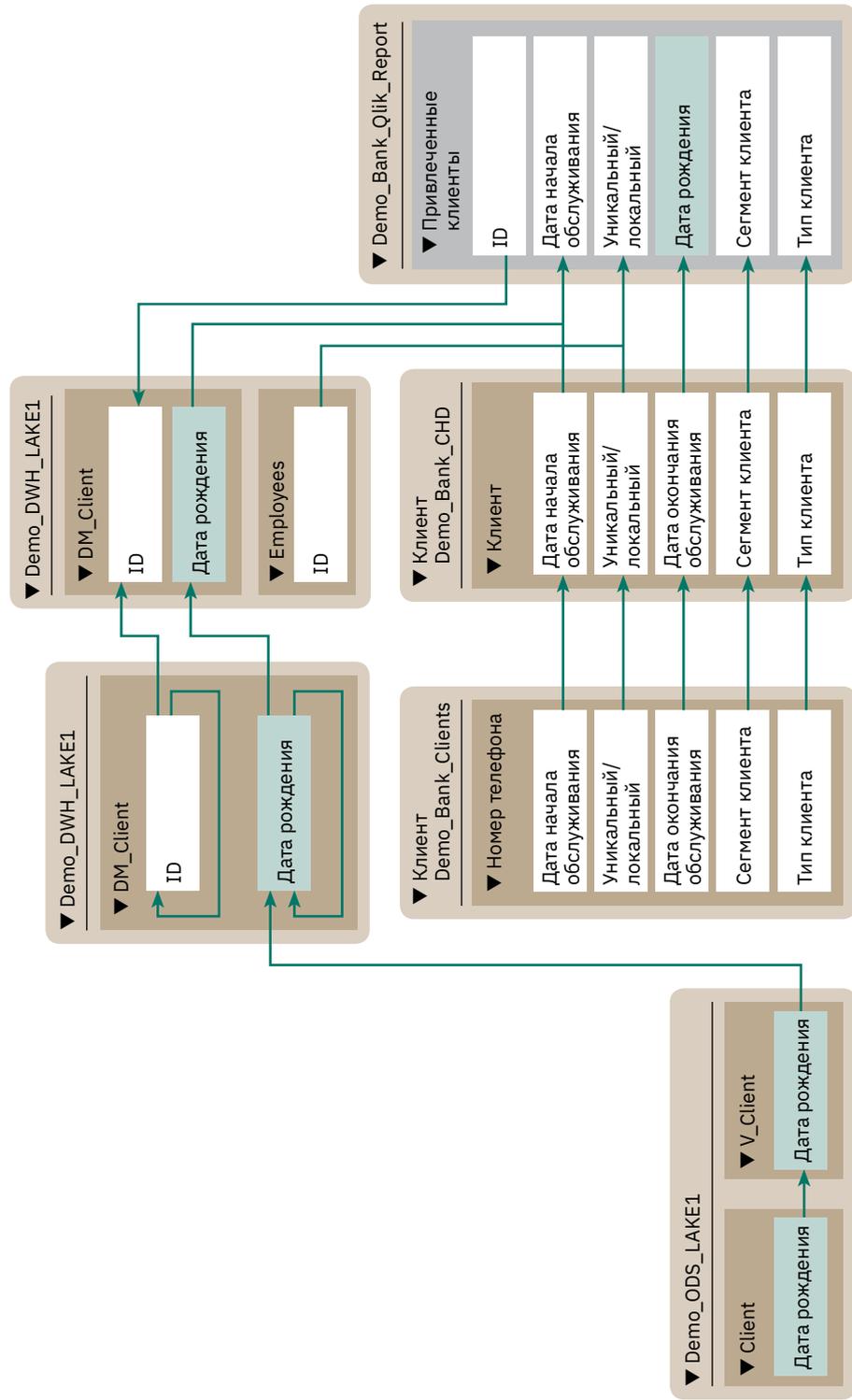


Рисунок 29. Пример детального представления Data Lineage — взаимосвязей данных организации в системах хранения и передачи информации

нужны для приоритетных задач организации. Бизнес-гlossарий и каталог данных останутся важнейшими инструментами управления данными и их качеством; регулярный аудит тоже должен стать частью этого процесса.

7.3 УПРАВЛЕНИЕ КАЧЕСТВОМ ДАННЫХ

Чтобы определить качество данных с разных позиций и в разных ситуациях, используются метрики качества данных: это основа управления качеством данных и самими данными¹⁵⁵. С развитием процессов ЦТ появилось множество метрик качества данных^{156, 157}.

Как правило, выделяют следующие метрики:

1) **полнота** — включают ли данные (и в какой степени) информацию, необходимую для конкретного бизнес-процесса;



Например, есть набор данных, связанный со спецодеждой. Чтобы обеспечить корректный процесс выдачи и замены спецодежды, нам нужны данные о человеке: вес, рост, размер ноги и т. д. В этом случае неполными будут данные, в которых нет размера ноги, и данные о 47 сотрудниках из 50.

2) **своевременность и волатильность** — описывают ли данные события в требуемом временном интервале;

3) **точность** — полное соответствие поставленным требованиям по значениям и заполнению;

4) **валидность** — соответствие данных заданному уровню достоверности, то есть корректно ли данные описывают объекты (неотрицательный вес, правильный формат даты, дата рождения в пределах последних ста лет). Валидные данные всегда надежны (точны);

5) **согласованность** — взаимная непротиворечивость данных;

6) **наличие и доступность** данных; под доступностью понимается время и усилия, которые требуются для получения данных в необходимом формате.

Существуют и другие метрики качества, разработанные для конкретных сфер деятельности организаций¹⁵⁸.

¹⁵⁵ Zhang R., Indulska M. & Sadiq S. Discovering Data Quality Problems // Business & Information Systems Engineering. 61. 2019. URL: https://www.researchgate.net/publication/334614038_Discovering_Data_Quality_Problems_The_Case_of_Repurposed_Data

¹⁵⁶ Jayawardene V., Sadiq S., Indulska M. An analysis of data quality dimensions. School of Information Technology and Electrical Engineering, The University of Queensland, ITEE Technical Report, 2013. URL: <https://espace.library.uq.edu.au/view/UQ:312314>

¹⁵⁷ Batini C., Cappiello C., Francalanci C., Maurino A. Methodologies for data quality assessment and improvement // ACM Computing Surveys. 41, 3. Article 16. 2009. URL: <http://doi.acm.org/10.1145/1541880.1541883>

¹⁵⁸ Например, иногда выделяют такой признак, как актуальность, указывающий на то, что на момент анализа данные не устарели и не появились новые данные.



Для создания и применения моделей количественной оценки кредитного риска¹⁵⁹ Центральный банк РФ применяет еще и такие метрики, как контролируемость и восстанавливаемость данных. Контролируемость данных — возможность осуществления контроля качества и происхождения данных, в том числе посредством отражения в ИС источников данных, истории создания, преобразования, изменения, удаления, хранения и передачи данных. Восстанавливаемость данных — возможность сохранять установленный уровень функциональности и качества данных после их утраты, повреждения или изменения в результате сбоев или иных нарушений работы ИС, ошибок или иных непредусмотренных действий персонала.

Управление качеством данных необходимо начинать с разработки методик и порядков обеспечения качества данных. Это позволит унифицировать процесс оценки качества данных. Ответственность за подготовку и своевременную актуализацию таких документов ложится на специалиста по качеству данных (см. раздел 5.1).

Методики обеспечения качества данных регламентируют:

- › показатели качества данных;
- › правила и методы измерения таких показателей;
- › показатели эффективности используемых методов и инструментов;
- › классификацию возможных источников и причин образования некачественных данных в системах;
- › критерии оценки качества с указанием пиковых значений всех показателей.

Порядки обеспечения качества данных регламентируют:

- › процедуры измерения показателей, указанных в методиках;
- › процедуры согласования и корректировки пиковых значений показателей;
- › процедуры реагирования на случаи несоблюдения критериев оценки качества данных и пиковых значений показателей;
- › процедуры корректировки ошибок и документирования внесенных изменений;
- › периодичность сверки данных и аудита качества;
- › порядок и периодичность создания отчетов о качестве данных;
- › порядок взаимодействия по вопросам обеспечения качества данных (полномочия, ответственность и обеспечение ресурсами, определение должностных лиц, несущих персональную ответственность за обеспечение качества данных).

¹⁵⁹ Центральный банк Российской Федерации. Положение от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов». URL: <https://docs.cntd.ru/document/420295268>



Рисунок 30. Структура документации, обеспечивающая эффективное управление качеством данных

Основные элементы систем, обеспечивающих описанные процессы, следует задокументировать, например в соответствии с требованиями Единой системы программной документации (ЕСПД). Структура такой документации должна иметь иерархический вид, как показано на рис. 30.

Оценку текущего уровня качества данных и выстраивание процессов управления качеством следует начать с аудиторской проверки документации на предмет соответствия требованиям, утверждения перечня лиц, ответственных за качество данных на административном и техническом уровнях, и определения перечня существующих правил качества данных.



Правительство Великобритании рекомендует использовать фреймворк «Качество государственных данных»¹⁶⁰.

Его первая часть — это структурные элементы, которые помогут организовать работу с качеством данных:

- › **принципы качества данных позволят создать культуру качества данных в организации;**
- › **руководство по жизненному циклу данных поможет организациям идентифицировать проблемы с качеством данных на всех этапах и смягчить их последствия;**
- › **метрики качества данных позволят проводить регулярную оценку качества данных.**

¹⁶⁰ Фреймворк был опубликован на сайте правительства Великобритании 3 декабря 2020 года, он состоит из двух частей и кейсов. URL: <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>

Вторая часть фреймворка — это практические инструменты и техники, которые применяются для оценки качества данных и его повышения:

- › **план действий** позволяет определить конкретные меры для оценки качества данных;
- › **анализ основных причин** позволяет решить изначальные проблемы, влияющие на качество данных;
- › **руководство по использованию метаданных** помогает более эффективно использовать метаданные для работы с качеством;
- › **руководство по коммуникации** (вопросы качества данных) содержит некоторые стратегии коммуникации с пользователями по вопросам качества данных;
- › **модели зрелости данных** позволяют специалистам применять целостный подход к оценке данных и повышению их качества.

Управление качеством данных не ограничивается разработкой методик и порядков. Важный шаг — внедрение программных решений (в том числе продуктов типа Data Quality¹⁶¹, см. табл. 4), которые соответствуют тем или иным метрикам качества.

Программные решения для управления качеством данных:

- › подробно анализируют качество данных, включая поиск связей между данными;
- › обеспечивают профилирование данных (выявляют статические характеристики данных);
- › находят и удаляют повторяющиеся записи;
- › стандартизируют (нормализуют) данные, то есть приводят их к эталонному унифицированному виду;
- › формируют отчеты о качестве данных.

Таблица 4. Программные решения, которые обеспечивают соответствие качества данных установленным требованиям

Метрика	Решение
Точность, полнота, валидность, волатильность	Классические продукты класса Data Quality
Согласованность и контролируемость данных	Программные продукты для ведения бизнес-гlossария и каталога данных/метаданных
Своевременность, доступность и восстанавливаемость данных	С помощью продуктов класса Data Quality можно осуществить проверку; качество по этим метрикам обеспечивается внутренними регламентами и процессами, а также инфраструктурными решениями

¹⁶¹ Data Quality, DQ (англ.) — качество данных.

7.4 ПАМЯТКА ДЛЯ ОТВЕТСТВЕННОГО ЗА КАЧЕСТВО ДАННЫХ

Некачественные данные подрывают веру в ценность данных. К снижению качества данных приводят неумелое планирование, неполнота документации, архитектурная обособленность систем, рассогласованность проектов. Работа с качеством данных должна быть встроена во все процессы организации.

1. Данные должны работать. Чем чаще данные используют, тем больше вероятность выявить ошибку (например, что в записи о гражданине правильная фамилия не «Фет», а «Фёт»).

2. Обеспечивать качество данных нужно на всех этапах процесса сбора или создания этих данных. Не должно быть таких ситуаций, в которых вся работа при сборе/получении данных ориентирована на их качество, но остается дополнительное «окно» для получения данных. К примеру, предоставить данные можно через электронную форму с масками ввода¹⁶² или в бумажном виде, где такой маски ввода нет. Другой пример: правила контроля качества запускаются автоматически, но на одном из этапов эта функция отдана оператору. При таком «слабом звене» все остальные усилия по обеспечению качества сильно обесцениваются.

3. Жизненный цикл данных должен быть встроен во все ключевые процессы организации, чтобы качество данных реально повышалось.

4. Человеческий фактор при сборе и вводе данных должен быть минимизирован. По всем метрикам должны быть внедрены форматно-логические контроли (см. раздел 3.2.1), различные маски ввода.

5. Ответственных за предоставление данных нужно мотивировать к тому, чтобы они вводили качественные данные. Например, если сотруднику при вводе информации выпадает только сообщение о том, что данные неполны, то он может не заполнить нужное поле или заполнить его некачественной информацией. А если показать ему сообщение, что отсутствие данных в конкретном поле приведет к неправильному начислению налога, шансы, что сотрудник заполнит поле правильно, возрастут.

6. Неправильная интерпретация данных может привести к социально значимым последствиям. Чтобы исключить неправильное понимание, данные должны быть описаны.

7. Необходимо всегда опираться на первоисточник. При передаче данных возможны ошибки, которые будут иметь критическое значение для получателя. Поэтому необходимо, чтобы всегда был указан первоисточник: данные должны максимально сохранять вид из первоисточника, а посредники должны их обогащать, а не заменять.

¹⁶² Маска ввода — строка символов, указывающая формат допустимых значений входных данных. Ее используют для обеспечения единого формата данных, например для полей с телефонными номерами из девяти цифр.



8. БЕЗОПАСНОСТЬ И РИСКИ ПРИ РАБОТЕ С ДАННЫМИ

 *Время чтения: 31 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Три главные угрозы данным — это утечка (разглашение), искажение, удаление (потеря). Степень критичности угроз каждая организация определяет для себя сама.
- › Если организация предоставляет гражданам цифровой сервис, то следует привлечь специалиста по информационной безопасности уже на этапе проектирования архитектуры сервиса, провести верхнеуровневый аудит архитектуры, сформировать карту функциональных ролей и на ее основе создать пакет средств и мер защиты.
- › Данные имеют уникальные характеристики, с которыми связаны и риски: данные неосязаемы, неисчерпаемы, их можно копировать, но сложно восстанавливать; данные порождают новые данные и могут использоваться многократно.
- › «Феодализм данных» — ограничение доступа к данным, которые хранятся в закрытых «колодцах» в организации. Он создает суперриск для организации: замедляет процессы и снижает ценность данных.
- › Информационная безопасность не сводится к разовым мерам или техническим инструментам, она подразумевает процесс, который следует организовать и поддерживать.

8.1 ЧТО УГРОЖАЕТ ДАННЫМ

Согласно DAMA-DMBOK, основным драйвером деятельности в области обеспечения информационной безопасности (ИБ) выступает стремление минимизировать риски и обеспечить устойчивое развитие организации¹⁶³. Для государственных организаций риски возникают еще и тогда, когда они нарушают свои юридические и моральные обязательства перед гражданами (а также сотрудниками и другими организациями), не обеспечив защиту личных сведений, конфиденциальных данных и прочей чувствительной информации. Утечки данных могут нанести репутационный ущерб государственной организации и отрасли в целом, подорвать доверие к ним.

Прежде чем сформулировать основные правила безопасной работы, опишем три главные угрозы данным: утечку (разглашение), искажение, удаление (потерю) данных. Степень критичности той или иной угрозы каждая организация должна определить для себя сама: событие, которое одной организации важно не допустить любой ценой, для другой может оказаться несущественным. На степень критичности той или иной угрозы для конкретной организации влияет критичность затрагиваемых ей данных.

КРИТИЧНОСТЬ (чувствительность) данных — важность данных, ценность, которая определяется субъектом, обрабатывающим данные. Одни и те же данные могут быть чувствительными для одной организации или сервиса и не критичными для другой.

1. Утечка (разглашение). Это самая распространенная угроза — регулярно появляются новости об утечках баз с ПДн пользователей. Однако сама по себе база необработанных данных интереса для злоумышленника чаще всего не представляет. Другое дело, что обычным явлением стало обогащение данных, когда несколько баз объединяются в одну с уже достаточно полной информацией о человеке (включая такую чувствительную информацию, как паспортные данные, наличие несовершеннолетних детей, медицинские данные и т. д.)¹⁶⁴.

Например, сервисы «Яндекса» аккумулируют не только поисковые предпочтения пользователя, но и, возможно, его передвижения («Карты»), данные о его питании («Еда»), покупках («Маркет»), коммуникациях («Почта»), досуге («Афиша»). Так создается полноценный цифровой отпечаток пользователя, цифровой двойник, «утечка» которого уже в недалеком будущем может иметь для человека катастрофические последствия.

Авторы главы:



А. В. Архипов



Е. А. Бойцова



И. М. Добровольский



Д. В. Ушаков

¹⁶³ DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп-Бизнес, 2020.

¹⁶⁴ О проблемах приватности и безопасности персональных данных см. серию аналитических докладов Центра руководителей и команд цифровой трансформации ВШГУ РАНХиГС: Этика и «цифра»: Этические проблемы цифровых технологий: в 2 т. М.: РАНХиГС, 2020; Этика и «цифра»: от проблем к решениям. М.: РАНХиГС, 2021. URL: ethics.cdto.center



В октябре 2021 года в Аргентине были похищены личные данные десятков миллионов граждан. Вначале в Сети появились данные 44 аргентинских знаменитостей, в том числе президента страны. Хакер, опубликовавший данные, заявил журналистам, что в его распоряжении есть полные имена, домашние адреса, даты рождения, дата выдачи и истечения срока действия удостоверений личности всех граждан Аргентины. Все эти данные хранились в национальном реестре граждан RENAPER, который используется для выпуска ID-карт¹⁶⁵.

2. Искажение. Чем сложнее модель данных, тем легче нарушить ее целостность, вследствие чего данные потеряют первоначальную ценность, а их владельцу (например, гражданину, цифровой отпечаток которого изменили злоумышленники) придется разбираться с драматическими последствиями искажений. К этому типу угроз относится и изменение конфиденциальной информации с целью влиять на экономическую активность организации. Крайняя степень искажения — шифрование данных программами, которыми злоумышленники полностью ограничивают доступ к данным и делают экономическую активность невозможной.

3. Удаление. Потеря данных, связанных с важными контрактами, чревата репутационными и финансовыми издержками и может влиять на способность организации продолжать работу. Потенциально чрезвычайно опасна потеря данных, представляющих собой юридически значимые записи в реестрах, начиная от ПДн граждан до записей об имуществе и т. д., особенно учитывая перспективу отмены бумажных документов. Разновидностью удаления данных может быть потеря их доступности в результате сбоев или злонамеренных действий (например, DDoS-атак¹⁶⁶).

На первый взгляд, самый простой способ решения перечисленных проблем — ограничить доступ к данным. Именно поэтому существуют закрытые ведомственные сети¹⁶⁷. Однако это далеко не всегда самый эффективный способ обеспечения безопасности данных при сохранении их ценности — ведь, например, для того чтобы аналитические агентства или спецслужбы могли получать ценные сведения или предотвращать преступления (не только в киберпространстве), данные должны быть доступны и открыты. Доступность с одной стороны и безопасность с другой — баланс между ними специалисту по безопасности каждый раз при создании цифровых сервисов приходится искать заново.

Забота о безопасности данных не означает замалчивание проблем — напротив, именно открытое обсуждение инцидентов помогает обеспечить сохранность информации. Инциденты важно не только расследовать, но и предавать их огласке, информировать пользователей о причинах

¹⁶⁵ Хакер похитил базу данных с информацией о всех жителях Аргентины и разместил объявление о продаже данных граждан // VC.ru. URL: <https://vc.ru/offline/307577-haker-pohitil-bazu-dannyh-s-informaciyey-o-vseh-zhitelyah-argentyiny-i-razmestil-obyavlenie-o-prodazhe-dannyh-grazhdan>

¹⁶⁶ Distributed Denial Of Service — распределенная атака с целью довести систему до отказа в обслуживании.

¹⁶⁷ Например, сеть RSNet. URL: <http://www.gov.ru/main/rsnet/page541.html>

и ответных мерах. Неслучайно во многих крупнейших цифровых компаниях мира — Microsoft, Apple, Yahoo и Facebook — правилом хорошего тона стало использование программ, гарантирующих сторонним пользователям вознаграждение (программы bug bounty — «награда за ошибку») в случае обнаружения уязвимостей и эксплойтов¹⁶⁸. Хранить секреты становится невыгодно: в открытом диалоге исследователи могут вносить свой вклад в безопасность цифрового пространства, а производители — делать свои продукты и решения более защищенными¹⁶⁹.

8.2 СВОЙСТВА ДАННЫХ И ВИДЫ РИСКОВ

Данные как актив (об этом см. раздел 1.2) имеют уникальные характеристики: в отличие от материальных активов их, например, нельзя исчерпать, они не изнашиваются и легко копируются. С этими особенностями данных связаны и риски при работе с ними.

1. Данные неосязаемы. Материальные активы можно потрогать, они находятся в определенном месте. Данные — нематериальный и неосязаемый актив, поэтому их характеристики сложно оценить.

Риски: несерьезное отношение к данным, отказ воспринимать их как актив, недостаточное внимание к данным, снижение их качества. При этом некачественные данные несут серьезную угрозу, поскольку приводят к неверным управленческим решениям.

Способы снижения риска:

- › формирование процедур и метрик для контроля качества данных (см. разделы 7.2, 7.3);
- › организация службы контроля качества данных;
- › проверка одних одних датасетов другими датасетами и иногда внешними данными (кросс-валидация);
- › просветительская работа, разъясняющая важность данных и их особые свойства как актива.

2. Данные неисчерпаемы, они не расходуются при использовании, в отличие от материальных и от некоторых видов нематериальных активов (например, лицензий).

Риск: на управление данными не выделяется достаточно средств. Возникает ложное впечатление, что управление данными не требует особых затрат. CDO или другому ответственному сотруднику непросто обосновать выделение средств — нужно понимание особенностей данных на уровне руководства организации.

¹⁶⁸ Эксплойты — подвид вредоносных программ, которые содержат данные или исполняемый код, способный воспользоваться уязвимостями в программном обеспечении на локальном или удаленном компьютере. См.: Что такое эксплойты и почему их все так боятся? // Kaspersky Daily. URL: <https://www.kaspersky.ru/blog/exploits-problem-explanation/8459/>

¹⁶⁹ Иногда замалчивание информации об инцидентах ведет к нежелательным последствиям для самой компании — например, в соответствии с Европейским законодательством о защите персональных данных (GDPR) компания может грозить кратный «оборотный» штраф, если она не разглашает вовремя информацию об инциденте.

Способы снижения риска:

- › выделение отдельного бюджета на управление данными;
- › создание организационной единицы для управления данными (см. об этом раздел 5.1);
- › проработка риска на уровне комитета по данным.

3. Данные имеют ценность, которую можно рассматривать в широком смысле как важность, значимость или полезность данных для организации и потребителей. Ценность данных не зависит напрямую от их возраста: они не могут испортиться, как земля, здания или оборудование. Однако их ценность со временем может меняться, причем в обе стороны.

Риск: обесценивание. Эксплуатация обесцененных данных приводит к ненужным тратам финансовых и трудовых ресурсов.

Способы снижения риска. Данные, утратившие свою ценность, следует выводить из активного обращения, архивировать и не использовать при наличии других, более ценных. Универсальной методики переоценки данных как актива не существует; ее следует проводить в ходе регулярного анализа процессов в организации. Иными словами, анализируя процессы на предмет их эффективности или соответствия процессной модели, следует задавать вопрос: насколько необходимы те или иные потоки данных в процессе?

Необходим регулярный пересмотр ценности наборов данных, а именно:

- › в случае повышения ценности — придание более высокого приоритета таким наборам данных при принятии управленческих решений, перемещение их в более быстрые хранилища, усиление защиты и т. п.;
- › в случае снижения ценности (например, при накоплении большого массива неиспользуемых данных) — принятие решения о выводе данных из эксплуатации и их архивации;
- › регулярный анализ и аудит существующих процессов в компании.

4. Данные можно копировать (реплицировать) без потери ценности и использовать для различных целей. Множество людей могут одновременно использовать одни и те же данные, что невозможно с материальными активами или финансовыми ресурсами.

Риск: репликация неверных данных или некорректная трактовка данных после репликации могут умножать ошибки. Недостаточно описанные данные могут неправильно трактоваться или неправильно использоваться после копирования. Не обновленные вовремя данные также могут приводить к ошибкам в их использовании. Копирование ошибочных данных может привести к экспоненциальному росту управленческих ошибок.

Способы снижения риска: создание паспортов для наборов данных с их детальным описанием.

5. Данные порождают новые данные; это верно для анализа данных, для выявления новых признаков при обогащении одних данных другими и т. д.

Риск: рост количества ошибок. Некоторые оперативные данные требуют дополнительной проверки (валидации другими данными); такие данные часто называют «горячими». Несвоевременное, раннее использование невалидированных данных может привести к росту ошибок. При интеграции бывает и обратная ситуация: в исходных системах в момент интеграции может не быть нужных наборов данных. Фактически при извлечении новых признаков или использовании «горячих» данных без должного контроля качества данных и регламентации их использования появляется риск роста количества ошибок в данных. Отдельно следует обратить внимание на то, что из нескольких наборов данных, по отдельности не содержащих ПДн, коммерческую или государственную тайну, при объединении и анализе иногда можно извлечь такую чувствительную информацию.

Способы снижения риска:

- › паспортизация наборов данных с отражением возможных новых признаков, которые могут быть выявлены при анализе;
- › определение порогов качества (толерантности) в паспортах, сигнализация при недостижении порогов качества;
- › использование моделирования для проверки качества данных;
- › контроль качества данных (см. раздел 7.3).

6. Данные сложно восстанавливать. Их легко скопировать и отправить куда угодно, но в случае уничтожения восстановить их будет крайне затруднительно.

Риск: потеря данных. Потеря данных может привести к приостановке деятельности, а в некоторых случаях и к катастрофам.

Способы снижения риска:

- › резервное копирование данных;
- › повышение отказоустойчивости и катастрофоустойчивости ИС.

7. Данные используются многократно. Данные не расходуются при использовании, поэтому их можно использовать многократно.

Риск: утечка. Хищение данных обычно незаметно, так как использование или копирование данных (см. риск 4) часто не оставляет никаких следов.

Способы снижения риска:

- › создание ролевой модели доступа к данным (см. раздел 8.3);
- › ведение журналов доступа к информационным системам;
- › применение специальных инструментов контроля утечки данных (Data Loss Prevention, DLP).

Суперриск: «феодализм данных» (см. раздел 4.5). В закрытых «колодцах данных» в организации, доступ к которым излишне ограничен, данные не только существенно снижают свою ценность, замедляют процессы, но и не позволяют увеличивать ценность, приумножать данные. При создании ролевой модели в организации должен быть использован принцип максимально возможной открытости наборов данных.

8.3 ДЕСЯТЬ ШАГОВ К ПОВЫШЕНИЮ БЕЗОПАСНОСТИ

ИБ не сводится к разовым мерам или техническим инструментам, она подразумевает процесс, который следует организовать и поддерживать. Если организация предоставляет гражданам цифровой сервис, то лучшее решение — привлечь квалифицированного специалиста по ИБ не на этапе эксплуатации, а в момент проектирования архитектуры сервиса, чтобы планомерно ставить и решать задачи, провести верхнеуровневый аудит архитектуры, сформировать карту функциональных ролей и на ее основе создать пакет конкретных средств и мер защиты.

Вот десять шагов в этом направлении, сформулированных в виде вопросов. Ответить на них владельцу сервиса стоит как можно раньше.

1. Что является критичным событием?

Прежде чем определять критичные события, можно на уровне анализа системы задать уточняющие вопросы: где расположены компоненты (объекты), с которыми работает система, и данные, которые она обрабатывает, кто является пользователем системы (субъектом)? Из ответов можно строить модель ИС, а также модель ее угроз и нарушителей. Когда модель сформирована и получен реестр недопустимых событий, можно оперировать уже более конкретными понятиями, думать о техниках и тактиках, которые используют злоумышленники, обращаться к базам угроз и уязвимостей Федеральной службы по техническому и экспортному контролю России (ФТЭК)¹⁷⁰ или MITRE¹⁷¹. Следующим этапом будет согласование и утверждение модели угроз у регулятора.



Предположим, перед командой стоит задача запустить цифровой сервис записи в поликлинику. Руководитель хочет сделать его безопасным. С чего начать? Есть законодательство о защите ПДн, требования в области объектов критичной информационной инфраструктуры (КИИ)¹⁷², но что именно необходимо делать — поставить межсетевой экран, установить безопасную операционную систему, избавляться от уязвимостей? И то, и другое, и третье приведет к повышению безопасности, но прежде всего следует разобраться, что руководитель считает критичным событием, которое никогда не должно произойти. Допустим, для сервиса записи в поликлинику важно, чтобы он работал 24/7, а данные пациентов не были искажены, удалены или украдены. Значит, критичным будет нарушение любого из этих сценариев. Из технических мер, кроме шифрования, нужны надежная аутентификация и разграничение доступа.

¹⁷⁰ См.: Банк данных угроз безопасности информации (БДУ ФТЭК). URL: bdu.fstec.ru/threat

¹⁷¹ MITRE ATT&CK® — база данных техник и тактик, привязанных к этапам реализации направленных атак, разработанная американской некоммерческой организацией MITRE. URL: <https://attack.mitre.org>
С 1999 года MITRE ведет также базу уязвимостей в программном обеспечении Common Vulnerabilities and Exposures. URL: <https://www.cve.org>

¹⁷² См. нормативную документацию с требованиями к безопасности Федеральной службы по техническому и экспортному контролю Российской Федерации (ФСТЭК России). URL: <https://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-spetsialnye-normativnye-dokumenty>

2. Сформирована ли команда специалистов по безопасности?

«Аутентификация», «авторизация», «шифрование», «обнаружение вторжений» — знакомые многим термины, но только специалисты¹⁷³ знают, как применять соответствующие средства защиты и выстраивать вокруг них процессы управления. ИБ не может быть гарантирована: завтра кто-то придумает новую атаку или найдет новую уязвимость. Защиту нельзя полностью автоматизировать, поэтому критически важные сервисы защищают живые команды в режиме 24/7.



Для совершенствования работы команды ИБ и отработки реакции на инциденты проводятся киберучения. В крупных компаниях они обычно проходят по следующему сценарию. Участники договариваются о тех угрозах, которые должны быть реализованы командой атакующих («красными», red team).

Команда безопасности компании («синие», blue team) должна их обнаружить. Противостояние развертывается в режиме реального времени. «Атакующие» должны незаметно пройти по инфраструктуре и «нажать красную кнопку». А «синие» — увидеть эти события и далее или заблокировать «красных», или — если атаке дают возможность развиваться — отследить ее на каждом шаге.

Команда ИБ работает хорошо, если, несмотря на срабатывания средств защиты, дело не доходит до инцидента безопасности (тем более до реализации рисков или нежелательных событий). Это означает, что:

- › корректно выбраны, установлены и настроены средства защиты;
- › написаны и исполнены регламенты выявления инцидентов безопасности и реагирования на них;
- › контроль безопасности настроен так, чтобы не допускать реализации рисков и пресекать вредоносную активность до ее приближения к критическому ресурсу.

Нельзя полагаться на то, что появление в команде архитектора по безопасности решит все проблемы. Предложенную модель угроз надо будет применить: установить и настроить средства защиты, создать систему управления информационной безопасностью (СУИБ). Для этого нужна команда из разных специалистов.

3. Внедрены ли средства защиты?

Огромное количество современных цифровых сервисов работает через веб-канал, поэтому необходимо обеспечивать безопасность соединения пользователя с сервисом с помощью шифрования канала. Это также дает возможность аутентификации по сертификату¹⁷⁴ на стороне сервера.

¹⁷³ В бизнесе верхнеуровневый специалист по безопасности обозначается аббревиатурой CISO — Chief Information Security Officer.

¹⁷⁴ По криптографическому протоколу, который подразумевает более безопасную связь. Сертификат безопасности SSL (Secure Sockets Layer) — это стандарт передачи данных с использованием цифровых подписей. Цифровая подпись сайта/системы/сервиса необходима для работы протокола защищенной передачи данных в Сети, поскольку именно она обеспечивает зашифрованное соединение между пользователем и сайтом.

На этапе выбора средств защиты команда специалистов проверяет:

- › Используются ли специальные межсетевые экраны уровня приложений — актуальна ли угроза взлома сервиса?
- › Используются ли подходы на основе безопасной разработки ПО?
- › Как реализована защита от DDoS-атак?
- › Как организован доступ к информации и работа с сервисом — через браузер или специально подготовленное клиентское приложение?

Если применены стандартные и проверенные протоколы, в них обнаружены и исправлены все основные уязвимости, то и сам сервис будет работать быстрее и стабильнее, чем полностью самостоятельная разработка. В выборе средств можно опираться на фреймворки некоммерческой организации The Center for Internet Security (CIS) — CIS Top Security Control¹⁷⁵.

4. Отделена ли открытая информация от конфиденциальной? Выполнена ли сегментация данных на уровне архитектуры сервиса?

Данные, файлы, базы данных, а также программная часть пользовательского интерфейса не должны находиться на одном сервере. Необходима сегментация и разнесение информации разных уровней конфиденциальности по разным объектам хранения и обработки. Объект из категории среднего уровня значимости не должен попадать сразу в категорию более высокого уровня значимости. Идентификаторы учетных записей пользователей, конкретные ФИО или паспортные данные или, тем более, номера медицинских карт — информация разных категорий. Соответственно, им требуется разный уровень защиты.

Лучшая практика — разносить данные разных категорий значимости по отдельным сегментам инфраструктуры и контролировать доступ субъектов (пользователей) из сегмента с одним уровнем значимости в сегмент с другим уровнем значимости. Разделение объектов с разным уровнем конфиденциальности позволяет не только упростить архитектуру, но и сделать ее более безопасной без нагромождения средств защиты.



В ВТБ создана математическая модель, которую можно использовать для привлечения клиентов, оценки рисков и персональных рекомендаций. С помощью методов машинного обучения последовательности действий клиента может быть представлена в виде обезличенных векторов (эмбеддингов¹⁷⁶), которые используются для предсказания поведения клиента. Векторная технология может применяться и для проактивного предоставления госуслуг. При этом она обеспечивает безопасность обработки данных: в ней нет ПДн.

¹⁷⁵ См. официальный сайт организации. URL: <https://www.cisecurity.org/controls/cis-controls-list>

¹⁷⁶ Эмбеддинги (от англ. embedding — вложение) — термин, который пришел в машинное обучение из технологий автоматической обработки естественного языка, где он обозначает процесс или, чаще, результат процесса преобразования по определенным правилам языковой сущности (слова, предложения или целого текста) в набор чисел — числовой вектор.

Когда данные должны быть доступны широкому кругу лиц, выше риск по недосмотру обнаружить то, что не предназначено для публикации. Комбинация данных также повышает уровень риска: при объединении нескольких наборов данных, не несущих конфиденциальной информации, можно вычлениить новые, неявные характеристики.

5. Обеспечен ли контроль доступа к данным?

Кроме сегментации данных, для разграничения доступа пользователей к разным сегментам используется такой базовый метод защиты данных, как аутентификация пользователя. Как она происходит?

1) Пароль. Для надежной защиты привычная пара «логин + пароль» не подходит, поэтому применяют методы двухфакторной аутентификации: в этом случае у пользователя есть второй фактор — программный или аппаратный токен, механизм синхронизации и оповещения.

2) Биометрическая аутентификация: отпечатки, голос, визуальные методы на основе анализа черт лица, микромоторики. Биометрические системы активно внедряются в банках. На уровне государства разработана Единая биометрическая система (ЕБС), которая использует комбинацию изображения и голоса¹⁷⁷. Обсуждая применение таких систем, эксперты указывают на сверхчувствительность биометрических данных: это «неизменные данные, которые достаточно украсть один раз навсегда»¹⁷⁸.

Для работы с документами разработаны методы аутентификации их источников и содержимого — электронно-цифровая подпись (ЭЦП). Раньше для ее использования был нужен физический носитель — карта памяти. Сейчас облачная ЭЦП может находиться в доверенном хранилище, и для доступа к ней достаточно смартфона и дополнительного фактора аутентификации при подписании документа (ОТР/Push-уведомления и СМС, которые приходят на доверенное устройство, или просто пароля).

6. Выделены ли разные категории пользователей?

С базами данных работают две категории пользователей: внутренние и внешние. У внутренних пользователей в целом больше прав по сравнению с внешними, но при этом у администратора полномочий значительно больше, чем у рядового пользователя системы (например, администратор может сразу скачать все записи из базы). Для разных категорий пользователей создаются разные «песочницы», или «капсулы», безопасности. Для внешнего пользователя достаточно входа по паролю и второму фактору аутентификации. Для внутреннего иногда создается особая программная среда, в которой у него в принципе нет возможности сохранить или выгрузить данные; в такой среде действует система контроля доступа, которая анализирует все виды активности пользователя.

¹⁷⁷ См.: Единая биометрическая система (ЕБС) // TAdviser. URL: [https://www.tadviser.ru/index.php/Продукт:Единая_биометрическая_система_\(ЕБС\)](https://www.tadviser.ru/index.php/Продукт:Единая_биометрическая_система_(ЕБС))

¹⁷⁸ Наталья Касперская: никаких особых способов защиты биометрии нет // РИА Новости. URL: <https://ria.ru/20211006/kasperskaya-1753227872.html>

Таблица 5. Уровни мониторинга безопасности данных

Уровень 1: осведомленность о жизнедеятельности систем	Уровень 2: сбор специальной информации	Уровень 3: аналитика на основе специальной информации
Все системы безопасности (межсетевые экраны, обнаружение вторжений и др.) и ИТ (серверы, базы данных) подключены к обычной системе мониторинга	Уровень 1 + ведение журналов событий безопасности сетевых устройств, узловых систем (агентов) с помощью системы сбора и обработки событий ИБ (SIM/SIEM).	Уровень 2 + метрики безопасности, которые обрабатываются аналитическими системами
Сбор атомарных событий	Собираются данные о данных. Есть отдельное облако данных, куда собирается информация о важных с точки зрения безопасности событиях	Получение информации о текущей защищенности объекта – Security Posture (многогранная метрика, привязанная к критичным бизнес-процессам и рискам и дающая объективную картину защищенности сервисов и систем)
Мониторинг непрерывности инфраструктуры и оборудования	Мониторинг отсутствия инцидентов безопасности	Мониторинг непрерывности процессов безопасности и состояния защищенности

7. Ведется ли мониторинг безопасности и управление системой защиты?

ИБ — не статический набор средств защиты и мер, а состояние, которое необходимо поддерживать. Регулярно появляются новые тактики атак, новые уязвимости. Приходят новые команды безопасности, беспечные сотрудники забывают сменить пароль. Причин для мониторинга много, вопрос в том, как его проводить. Иногда организации идут по формальному пути и начинают строить систему менеджмента ИБ по серии стандартов¹⁷⁹ ГОСТ Р ИСО/МЭК 2700х. Это правильный путь, но долгий, он касается комплексного управления безопасностью в компании. Если же рассматривать мониторинг как возможность в любой момент времени ответить на вопрос «Что сейчас происходит с системой (сервисом)?», то такую задачу можно решать на разных уровнях, описанных в табл. 5.

Настроить процессы безопасности — значит отнестись к данным как к активу и наладить управление средствами защиты, аудита, реагирования на инциденты в соответствии с уровнем развития инфраструктуры¹⁸⁰. Необходимо настроить аналитику для понимания состояния защищенности системы. Средства безопасности должны акцентировать внимание специалиста по безопасности на проблемах. Метрика, позволяющая однозначно утверждать, что команда безопасности и созданная ею система работают хорошо, для каждой организации формируется индивидуально.

Аудит уже работающего сервиса можно провести в виде тестирования на проникновение внутрь системы (англ. penetration test, «пентест»). Почти все пентесты заканчиваются проникновением, но у руководителя появляется

¹⁷⁹ Национальный стандарт Российской Федерации «Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности» // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/1200058325>

¹⁸⁰ Можно опереться на стандарты серии ГОСТ Р ИСО/МЭК 2700х.

понимание общей картины и проблем, которые стоит устранить в первую очередь (например, закрыть уязвимости или переписать веб-приложение).



В ходе международного аудита Организации Объединенных Наций по промышленному развитию (ЮНИДО) в начале 2021 года аудиторы Счетной палаты РФ провели тестирование на проникновение в ИТ-системы организации и сравнительный анализ готовности к кибератакам. Системы имели продвинутую «многослойную» защиту, но в некоторых случаях аудиторам удалось получить доступ к конфиденциальной информации: ПДн сотрудников, записям совещаний руководства организаций, к данным, хранящимся в ИТ-системах. Такой доступ дает возможность, к примеру, подменить банковские реквизиты, что в момент оплаты позволяет перевести деньги не поставщику, а злоумышленнику. Аудиторы также смогли получить доступ к отдельным файлам и целым папкам, где хранились ссылки на все файлы организации, которыми сотрудники когда-либо «делились». Многие организации системы ООН по результатам проведенных аудитов приняли стратегическое решение пересмотреть подходы и выделили в ИТ-службе сотрудника, который будет отвечать за локализацию политик ООН, а также выполнять мониторинг инцидентов и анализировать возникающие риски нарушения кибербезопасности.

8. Обучены ли сотрудники основам безопасной работы с данными?

Человеческий фактор — один из наиболее существенных источников угроз. Как бы хорошо сервис ни был защищен, уровень цифровой грамотности пользователей, которые работают с сервисом, часто невысок, и тогда они становятся «точкой» проникновения внутрь системы, жертвами социальной инженерии. Еще более опасен низкий уровень цифровой грамотности внутренних привилегированных пользователей, которые имеют право, например, обслуживать базу данных.

9. Проработана ли отказоустойчивость сервисов, введена ли практика резервного копирования данных? Если да, то минимизированы ли риски при обеспечении отказоустойчивости и резервном копировании?

Под отказоустойчивостью обычно понимают свойство сервиса сохранять работоспособность и продолжать выполнение задач, даже если отдельные компоненты ИТ-инфраструктуры выходят из строя. Предотвратить угрозу безвозвратной потери данных, сделать инфраструктуру более надежной позволяет практика резервного копирования (создания бэкапов).

10. Обеспечена ли безопасность процессов разработки (в случае, если организация располагает собственными разработчиками)?

Для надежной и безопасной разработки необходимо использовать классический ИТ-ландшафт: среду разработки, среду тестирования и продуктивную среду. Все три среды должны быть независимы друг от друга — данные между ними не должны передаваться. В процессе разработки сервиса для разработчика важно иметь наборы тестовых данных в тестовой среде, не несущие чувствительной информации.



9. АНАЛИТИКА: ПРЕОБРАЗОВАНИЕ ДАННЫХ В ЗНАНИЯ И РЕШЕНИЯ

 *Время чтения: 35 мин.*

КЛЮЧЕВЫЕ ИДЕИ

- › Аналитика помогает обосновать управленческие решения на четырех уровнях: описательном («Что произошло?»), диагностическом («Почему это произошло?»), предсказывающем («Что произойдет?») и предписывающем (когда руководители получают варианты решения проблемы).
- › Data Mining – направление аналитики, которое применяется к очень большим объемам данных, когда перечислить все гипотезы невозможно. К нему примыкает наука о данных, в частности, когнитивный анализ данных с помощью методов машинного обучения.
- › Создание развитой аналитики не всегда требует больших расходов. Организации могут начать с использования существующих продуктов.
- › Системы бизнес-аналитики позволяют быстро собрать информацию, обработать ее, визуализировать (в том числе в виде дашбордов) и сделать выводы. Они могут с успехом применяться и в сфере госуправления.
- › Современному руководителю важно наладить процессы таким образом, чтобы аналитики занимались только своей работой, без необходимости готовить данные, описывать их, собирать, обеспечивать качество.

9.1 БАЗОВЫЕ ОПЕРАЦИИ ИЗМЕРЕНИЯ

Авторы главы:



Я. Э. Гейн



Д. Г. Давыдов



И. М. Добровольский



А. В. Новичков



М. В. Туманова

Если данных немного, выводы из них сделать легко: например, ничего не стоит сравнить штатную численность сотрудников в двух или трех отделах. Но часто удержать все данные в уме не представляется возможным. Это и не нужно, поскольку обычно требуются не все сведения о процессах и явлениях, а только некоторые. Именно отсюда — само понятие «**анализ**» (от греч. analysis — разложение): это вычленение из изучаемой реальности отдельных черт, свойств или отношений. Аналитик взаимодействует не с реальностью, а с ее «следами». Такими «следами» становятся данные, которые естественно возникают в результате деятельности людей (например, сведения о численности населения или объеме финансирования проектов) или создаются специально (данные опросов).

С точки зрения аналитика все данные — это результат измерения. Согласно определению физиолога и психолога Стенли Смита Стивенса, **измерение** в широком смысле — это «приписывание числовых форм объектам или событиям в соответствии с правилами»¹⁸¹. К объектам физического или социального мира относятся люди, организации, территории, мнения и т. п. Правила приписывания формируют шкалу измерения. С полученными в результате приписывания значениями и работает аналитик, пытаясь построить полезную для практиков модель реальности.

Хотя измерение физических величин — это простая и очевидная операция, даже здесь необходимо понимать, каким способом проведено такое «приписывание». Если забыть об этом, один и тот же объект измерения может оказаться разной длины (когда персонажи мультфильма «38 попугаев» измеряли удава, «в попугаях» он оказался длиннее, чем «в мартышках»).



Иногда такие ошибки имеют нешуточные последствия: в 1999 году космический аппарат Mars Climate Orbiter стоимостью 125 млн долл. врезался в Марс. Оказалось, что часть программного обеспечения рассчитала силу двигателей в британской имперской системе мер, а не в метрической, как должна была¹⁸².

Аналитика данных базируется, среди прочего, на теории вероятностей и математической статистике, но развивалась она при

¹⁸¹ Stevens S. S. On the Theory of Scales of Measurement // Science. 7 Jun 1946. Vol. 103. Issue 2684. Pp. 677–680. URL: <https://www.science.org/lookup/doi/10.1126/science.103.2684.677> В то же время Стивенс выделяет и нечисловое приписывание — по шкале наименований. По Стивенсу, помимо нее есть еще три основных шкалы измерений: порядковая, интервальная, относительная. Только последняя имеет нулевую точку, подразумевающую действительное отсутствие измеряемого качества (например, длины), а значит, и возможность деления и умножения. Для интервальной шкалы — например, температурной — деление невозможно: нельзя утверждать, что 30 °C — это в два раза теплее, чем 15 °C (так как 0 °C — это не отсутствие температуры).

¹⁸² Mars Climate Orbiter сгорел, не успев выйти на работу // Новости космонавтики. № 11. 1999.

решении прикладных задач экономики, биологии, социологии, управления качеством. Тысячи алгоритмов можно свести к **пяти базовым операциям**.

1. Оценка уровня чего-либо (в абсолютных значениях), относительная оценка (например, доля от общего) и центральная тенденция (например, среднее арифметическое).
2. Оценка распределения частот (по объектам, категориям и т. д.).
3. Сравнение двух или более значений (например, уровней одного качества у разных объектов) или сдвиг (разница между текущим и предшествующим уровнями одного объекта). Аналитика здесь обычно предстает в виде коэффициентов, описывающих статистическую достоверность различий.
4. Оценка динамики: измерение состояния в течение периода времени.
5. Оценка взаимосвязи между двумя или более явлениями. Сюда относятся разнообразные специальные критерии (регрессия, коэффициенты корреляции, сопряженности и т. п.). Если одно явление меняется, будет ли меняться другое? Связана ли, например, доступность спортивных объектов с уровнем криминального или протестного поведения молодежи? А активность подразделения в организации — с тем, что у этого подразделения выше прибыль?



Руководителю, который работает с аналитикой, важно понимать, что корреляция между двумя явлениями не говорит о наличии причинно-следственной связи между ними. Корреляция может быть случайной — например, в США количество убийств с 2006 по 2011 год снижалось с той же скоростью, что и использование Microsoft Internet Explorer. Корреляция между двумя переменными может быть обусловлена третьей переменной. Так, число пожаров в населенном пункте коррелирует с числом медиков, работающих там. Но здесь нет прямой причинности: сократив число пожаров, мы не уменьшим число медиков. Численность населения — та неочевидная третья переменная, которая влияет на первые две. Так же устроена ложная корреляция между ростом продаж мороженого и увеличением числа случаев утопления (заставить людей чаще плавать, а также покупать больше мороженого может жаркая погода).

С помощью базовых операций можно решить большинство прикладных задач. Продвинутые аналитические алгоритмы также построены на базовых операциях (например, факторный анализ предполагает множественные оценки корреляций между набором переменных).

9.2 УРОВНИ И РЕЗУЛЬТАТЫ АНАЛИТИКИ

Типичные проекты в области анализа данных делятся на несколько этапов¹⁸³: данные оцениваются и отбираются, очищаются и фильтруются,

¹⁸³ Межотраслевой стандартный процесс интеллектуального анализа данных (Cross-Industry Standard Process for Data Mining, CRISP-DM) включает шесть этапов: понимание целей, начальное изучение данных, подготовка данных, моделирование, оценка, внедрение.

визуализируются и анализируются, а результаты анализа в итоге интерпретируются и оцениваются. Трудоемкость этапов неодинакова. Львиная доля рабочего времени (принято считать, что 80–90%) уходит на поиск, извлечение, очистку данных (см. разделы 6.2 и 4.1) и подготовку их к анализу¹⁸⁴. Для удобства можно выделить ключевые этапы аналитического проекта, показанные на рис. 31.

Аналитика помогает принимать обоснованные управленческие решения на четырех уровнях¹⁸⁵ (табл. 6), каждому из которых соответствует определенный результат.

Первый уровень — **дескриптивный (описательный)**, когда мы можем описать прошлое и ответить на вопрос «Что произошло?». Это, например, элементы бухгалтерского учета или количество тех или иных записей в ИС, число жителей в разных районах области, количество автодорог на душу населения. По оценкам авторов, на этом уровне находится около 70% организаций, как государственных, так и негосударственных. Для большинства повседневных задач описательного уровня достаточно.

Второй уровень — **диагностический**; он отвечает на вопрос «Почему что-либо произошло?». На нем появляется собственно анализ, например ключевые показатели, рентабельность и аналитический финансовый показатель EBITDA в бухгалтерской отчетности; динамика изменения фактического населения региона по сезонам года, изменение удовлетворенности населения после введения каких-либо новых услуг. На этом уровне находится 15–20% организаций госсектора.



Рисунок 31. Этапы аналитического проекта

Предиктивная (прогностическая, предсказывающая) аналитика¹⁸⁶ — это уже не просто анализ, но и построение различных моделей на основе данных, которыми располагает организация, — от простых регрессионных до более сложных на базе нейронных сетей. Предиктивная аналитика отвечает на вопрос «Что произойдет?».

¹⁸⁴ Перед этапом анализа выделяется также процесс самостоятельной подготовки данных (Self-Service Data Preparation), цель которого — привести данные разной периодичности, форматов и качества в единый вид силами бизнес-пользователей. Отличие от стандартного сбора (интеграции данных) заключается в том, что подготовку делает не инженер данных, а сам бизнес-пользователь.

¹⁸⁵ Four ways governments can use data to transform outcomes // EY.
URL: https://www.ey.com/en_gl/future-government/four-ways-governments-can-use-data-to-transform-outcomes

¹⁸⁶ О том, как работают предиктивные аналитические модели в современных организациях мира, см.: Сигель Э. Просчитать будущее. Кто кликнет, купит, соврет или умрет. М.: Альпина Бизнес Букс, 2014.



Центр ЦТ ФАУ «Главгосэкспертиза России» разрабатывает цифровые решения на основе дата-ориентированного подхода для отрасли строительной экспертизы. В ходе реализации нацпроекта «Цифровая экономика» создается Единая цифровая платформа экспертизы. В нее загружается в машиночитаемом виде вся предоставляемая на экспертизу документация, производится разметка данных, что позволяет применять аналитические методы и получать ответы на вопросы «Почему?», «Кто?» и «Какие действия предпринять?»¹⁸⁷. Возможности цифровой платформы экономят время и средства при проведении экспертизы.

Наконец, на **прескриптивном (предписывающем)** уровне руководители получают варианты решений анализируемой проблемы. Реализована такая аналитика может быть как в виде расчетов вероятности тех или иных событий при принятии определенного решения или поиска оптимума «затраты — доходы», так и в виде полноценных систем поддержки принятия решений. Лишь считанные проценты организаций в мире строят свою деятельность на предиктивной или прескриптивной аналитике.



«Издавек аналитические информационные панели могут выглядеть внушительно: какие-то цифры, графики. Но представьте, что это панель управления автомобиля — куча мелких цифр и диаграмм. Как вы будете им управлять?»

Хорошая аналитическая система должна включать прогноз, а не одни мертвые данные, показывающие, что произошло. Если у вас нет прогноза, вы не понимаете, как будет развиваться ситуация».

Илия Димитров, омбудсмен по вопросам развития цифровой экономики

Следует учесть, что системы предиктивной или прескриптивной аналитики нуждаются в больших массивах данных, создание которых требует отдельных усилий и специальной инфраструктуры. Такого уровня аналитики легче достичь, если организации удастся преодолеть «феодализм данных» (см. также раздел 4.5), когда данные хранятся изолированно в локальных базах отделов и департаментов, а их интеграция в единые наборы затруднена по административным или техническим причинам.

Развитию высокоуровневой аналитики способствует свободный обмен данными, накопление их в специальных хранилищах, постоянное обогащение данных. Кроме того, важно наличие метаданных (сведений об имеющихся данных), регламентов сбора, обработки и хранения данных.

¹⁸⁷ BIG DATA 2020: Главгосэкспертиза на единой цифровой платформе // Вестник цифровой трансформации CIO.RU. URL: <https://www.computerworld.ru/cio/articles/090320-BIG-DATA-2020-Glavgosexpertiza-na-edinoy-tsifrovoy-platforme>

Таблица 6. Уровни аналитики данных

Уровень	Дескриптивный	Диагностический	Предиктивный	Прескриптивный
Основной вопрос	Что случилось? Что происходит?	Почему это произошло?	Что, скорее всего, произойдет в будущем?	Какой вариант самый лучший и как его достичь?
Основной фокус	Анализ и визуализация данных для понимания того, что уже произошло или происходит в режиме реального времени	Анализ исторических данных с целью обнаружения причинно-следственных связей, закономерностей и корреляций	Моделирование тенденций и возможных изменений на основе данных о том, что уже произошло, и разработка гипотез о том, что может произойти в будущем	Комплексная аналитика с целью определения наиболее благоприятного сценария и анализ способов его реализовать
Применение в госсекторе	Отчетность, ИС, демографическое профилирование, мониторинг и оценка	Изучение причинно-следственных связей, таких как влияние экономических факторов на уровень преступности или связь образования с доходом и экономическим ростом	Разработка сценариев, стратегическое планирование и моделирование (прогнозы экономического роста, скорость распространения болезней, а также иммиграция и рост населения)	Поддержка принятия решений в таких областях, как социальные услуги и правосудие; обоснования для долгосрочных инвестиций
Примеры использования	«Умные» города в режиме реального времени анализируют потребление энергии и корректируют его в общественных зданиях в часы пик. В Бостоне такие системы экономят до 1 млн долл. в год	Министерство транспорта США инициировало программу SDI (Safety Data Initiative); она основана на анализе данных и выявляет точки повышенного риска аварий на дорогах и железно-дорожных путях	Прогностические модели для пенсионных фондов, показывающие, как экономические, политические и социальные изменения влияют на финансовое поведение	Выбор маршрутов для патрулирования полицейскими. Рекомендации по изменению маршрутов общественного транспорта, по развитию дорожной сети

9.3 ТРЕНДЫ АНАЛИЗА ДАННЫХ

Традиционная аналитика (оценка признака, различий, взаимосвязей и т. д.), описанная в разделе 9.1, остается востребованной; владение ее инструментами формирует мышление аналитиков и управленцев, позволяет принимать верные решения на основе данных. Вместе с тем в последние годы развиваются новые направления, в которых работают как аналитики, так и исследователи данных. Если аналитик получает из данных те или иные «следы», элементы реальности (например, динамику численности населения), то исследователь еще и извлекает данные, работает с неструктурированными данными; применяет математические модели для аналитических задач, прогнозирования, разработки на их основе приложений и сервисов.

Интеллектуальный анализ данных (Data Mining) — направление аналитики, которое применяется к очень большим объемам данных, когда вариантов их интерпретации слишком много и перечислить все гипотезы невозможно. Задача в таком случае — извлекать из данных полезные догадки (такого рода озарения называют «инсайтами»), черпать идеи, находить неожиданные взаимосвязи и скрытые закономерности.



Задачей для интеллектуального анализа данных может быть, например, сегментация пользователей на основе кластерного анализа, выявление стилей поведения, поиск аномалий в данных. Еще одна задача — определение сходства между объектами. Насколько два объекта похожи между собой по набору критериев?

Возможно ли, например, сгруппировать районы, области, населенные пункты по сходным проблемам или перспективам развития? Можно ли — если идти еще дальше — выявить кластеры, группы таких объектов? Как люди, так и объекты управления (отделы, города, регионы) всегда индивидуальны, но выявление общих признаков существенно улучшает качество и скорость работы с ними.

Когнитивный анализ данных с помощью методов **машинного обучения** (Machine Learning) развивается в **науке о данных** (Data Science) — области знания, тесно связанной с интеллектуальным анализом данных, предполагающей, как правило, использование больших массивов данных и мощных вычислительных средств. Для такого анализа разрабатываются алгоритмы, имитирующие некоторые когнитивные способности человека.

Предположим, нам надо изучить содержание тысяч электронных писем, которые получает организация, и комментарии к ее постам в соцсетях. Можно применить систему работы с текстами, которая проанализирует тип обращений, количество сообщений того или иного типа и даже разделит комментарии по «знаку» эмоциональной оценки на положительные, нейтральные, отрицательные.



«Некоторые из наших университетов-партнеров сообщили, что лучше всего продаются курсы по Data Science и продвинутой аналитике. И это здорово, это модно, это круто и весело. Но это не волшебная таблетка и не серебряная пуля. Работа этих технологий зависит от степени достоверности информации. Мы не можем просто поверить в популярную технологию, про которую люди говорят: „Возьмите свои данные и положите их в озеро данных, и случится волшебство“. Потому что это неправда, не так ли?»¹⁸⁸

**Кристофер Брэдли, вице-президент
по профессиональному развитию DAMA International**

¹⁸⁸ Иткин В. Интервью с Кристофером Брэдли — одним из авторов DAMA DMBOK // VC.ru. URL: <https://vc.ru/video/210338-intervyu-s-kristoferom-bredli-odnim-iz-avtorov-dama-dmbok>

9.4 СПОСОБЫ ВИЗУАЛИЗАЦИИ ОТ ГРАФИКА ДО ДАШБОРДА

Визуализация данных — еще одно мощное направление современного анализа данных. Визуализацией называют представление данных с помощью наглядных изображений (обычно — разнообразных диаграмм). Задача визуализации — не отобразить данные максимально точно (например, разницу в одну минуту или один метр), а показать, какие есть тенденции, помочь быстро понять, что больше, что меньше и как велико различие.

Визуальное представление данных имеет довольно длинную историю. Первые графики, отражающие в виде линий динамику экономических процессов, были известны уже в середине XVIII века. Сейчас появились новые инструменты: многомерные графики, облака тегов, тепловые карты, системы бизнес-аналитики (см. раздел 9.5.1).

Визуализация — это общий термин, который описывает как способность мозга создавать зрительные образы, так и техники материального воплощения этих образов. Чаще всего визуализация необходима там, где изучают и обсуждают абстрактные сущности, такие как схема организации, схема производства, динамика изменения показателей, сравнение категорий в рейтинге и т. д., и принимают решения на их основе.

Графические визуализации — это прежде всего блок-схемы и диаграммы. На **блок-схемах** показаны не столько данные, сколько процессы, связи: блоки, описывающие понятия, соединены линиями, обозначающими связь между ними. **Диаграммой** называют графическое построение в прямоугольной или радиальной системе координат. Их используют для визуального исследования данных, для результатов измерений, расположения во времени значений одного или нескольких рядов. В отличие от блок-схем, диаграммы трудно понять без пояснений («легенды») и дополнительных графических элементов (осей и сеток).

Универсальные принципы превращения обычных неясных графиков в мощные инструменты донесения идей можно узнать из книги Александра Богачева «Графики, которые убеждают всех» (АСТ, 2020).

«Любой аспект данных, на который вы хотите обратить внимание с помощью диаграммы, может быть выражен посредством одного из пяти основных типов сравнения», — утверждает Джин Желязны, автор пособия по визуальным коммуникациям для руководителей и аналитиков¹⁸⁹. Вот эти пять типов сравнения: покомпонентное (процент от целого), позиционное (расположение объектов, одинаковы ли они, больше или меньше других), временное (изменения во времени), частотное (объекты в интервалах),

¹⁸⁹ Желязны Д. Говори на языке диаграмм. М.: МИФ, 2011.

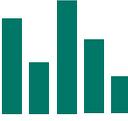
	Покомпонентное	Позиционное	Временное	Частотное	Корреляционное
Круговая					
Линейная					
Гистограмма					
График					
Точечная					

Рисунок 32. Виды диаграмм и типы сравнения, которые в них используются

корреляционное (зависимость между переменными). Каждый из этих типов сравнения Железны соотносит с одним из пяти видов диаграмм, показанных на рис. 32.

Существует множество инструментов автоматического создания схем и диаграмм (см. рис. 33). Для визуализации данных созданы специализированные технологии конструирования информационных панелей, хотя инструменты визуализации встроены и в офисные пакеты.

Облачные технологии создания диаграмм и визуализаций

1. Infogram
2. Datawrapper
3. Flourish
4. RAWGraphs
5. Data Illustrator
6. Charticulator
7. Miro

Технологии конструирования информационных панелей (дашбордов)

1. Tableau
2. Microsoft Power BI
3. Yandex DataLens
4. Google Data Studio

Программы графического редактирования

1. Adobe Illustrator
2. Adobe XD
3. Figma

Офисные пакеты документооборота

1. Google Docs
2. MS Office
3. Libre Office
4. Open Office

Дополнения для построения диаграмм

1. Chart
2. Datology



Рисунок 33. Визуализация: примеры технологий

Платформы для построения блок-схем и диаграмм доступны и в виде облачных решений, не требующих установки на компьютер пользователя.

Если необходимо представить данные массовой аудитории в маркетинговых или просветительских целях, то значимыми становятся идентификация с брендом и привлекательность на фоне повседневной визуальной коммуникации. Такого рода продукт принято называть информационной графикой, или, сокращенно, **инфографикой**. Для ее создания рекомендуется обращаться к профессионалам, которые помимо умения работать с данными обладают опытом в области графического дизайна¹⁹⁰.



«Среди лучших образцов современной отечественной инфографики – работы медиапроекта „Эконс“ (ЦБ России), проекты „Газпром нефти“ и студии „Инфографика“ для Европейского союза. Особого внимания заслуживают работы студии инфографики ТАСС, лидера в России как по числу наград, так и по количеству работ и просмотров».

Алексей Новичков, доцент факультета коммуникаций, медиа и дизайна НИУ ВШЭ

Эффективность визуализации данных в медиа привела к распространению этого формата презентации и хранения данных и в документообороте. Использование наглядных диаграмм в выступлениях и в отчетах стало стандартом качества коммуникаций. Помимо документооборота, визуализация используется в системах контроля показателей на интерактивных досках управления и дашбордах (см. раздел 9.5.1); в этом случае ценятся информативность и точность отображаемых данных.



Ошибки визуализации: три главных НЕ

- 1. Не рассчитывайте самостоятельно, не привлекая профессионалов, сделать хорошую визуализацию или инфографику¹⁹¹.**
- 2. Не выбирайте тип диаграммы, пока не определитесь с задачей. Сначала поймите, что хотите сделать, какие данные у вас есть, что это за данные и зачем они нужны.**
- 3. Не забывайте о базовых принципах восприятия информации. Со школьных времен ось X мы воспринимаем как ось изменения – и это должно быть упорядоченное изменение. Никогда не вбивайте в круговую диаграмму значения, в сумме не дающие 100%. Например, из одного отдела уволилось пять сотрудников, из другого – восемь, но вместе они не дают 100%!**

¹⁹⁰ Представление о работе профессионалов в области инфографики и дата-арта дает ежегодная международная премия Moscow Dataviz Awards. URL: <https://moscowdatavizawards.com>

¹⁹¹ Проверить свою грамотность в области визуализации данных можно с помощью сервиса bar-or-pie.dianov.org. Игровой тест показывает типичные случаи нарушения правил визуализации и помогает в этих правилах разобраться.

9.5 ПОСТРОЕНИЕ АНАЛИТИКИ В ОРГАНИЗАЦИИ

9.5.1 ВЫБОР ИНСТРУМЕНТА В ЗАВИСИМОСТИ ОТ ЗАДАЧИ

Создание развитой аналитики не всегда влечет большие расходы, хотя существуют дорогие многофункциональные промышленные аналитические системы, требующие высокой квалификации пользователя. Любой современный компьютер оснащен инструментами, позволяющими анализировать довольно большие массивы данных и получать нужные решения.

1. Офисные пакеты

Существующие комплексы офисных продуктов могут стать точкой входа для организаций, которые ранее не использовали аналитику и не имеют в штате обученных сотрудников. Однако «точка входа» не значит «примитивный инструмент». Электронные таблицы, входящие в такие комплексы (в том числе российские «МойОфис» и «Р7-офис»), можно подключать (скажем, в MS Excel) к разным источникам, работать с наборами данных до миллиона строк, выполнять серьезные преобразования, строить дашборды и т. д.

2. Аналитика, встроенная в программные продукты

Сегодня многие производители технических систем, ПО и цифровых сервисов предоставляют пользователям доступ ко встроенным аналитическим решениям. Так пользователи получают возможность анализировать работу самой системы и собираемые с ее помощью данные. Сложность, гибкость настройки и удобство подобных аналитических модулей весьма различны, но в любом случае при приобретении ПО или цифровых сервисов стоит учесть наличие встроенной аналитики.



Примеры встроенной аналитики:

- › счетчик для веб-сайтов Google Analytics помимо сбора данных строит дашборды, отображающие поведение пользователей;
- › системы управления образовательной средой (Learning Management Systems, LMS) имеют модуль аналитики, который позволяет оценить динамику прохождения курса, поведение обучающихся и т. п.;
- › медицинские системы выводят накопленные данные в виде обобщенной динамики, описательных статистик.

3. Комплексные универсальные статистические программы

Пакеты статистического анализа (IBM SPSS, программа STATISTICA компании StatSoft, SAS) позволяют работать с разными направлениями статистики и имеют несколько базовых функций: классическую статистику, построение визуализаций, Data Mining и типовые алгоритмы машинного обучения. У каждого такого пакета есть, как правило, свой язык.

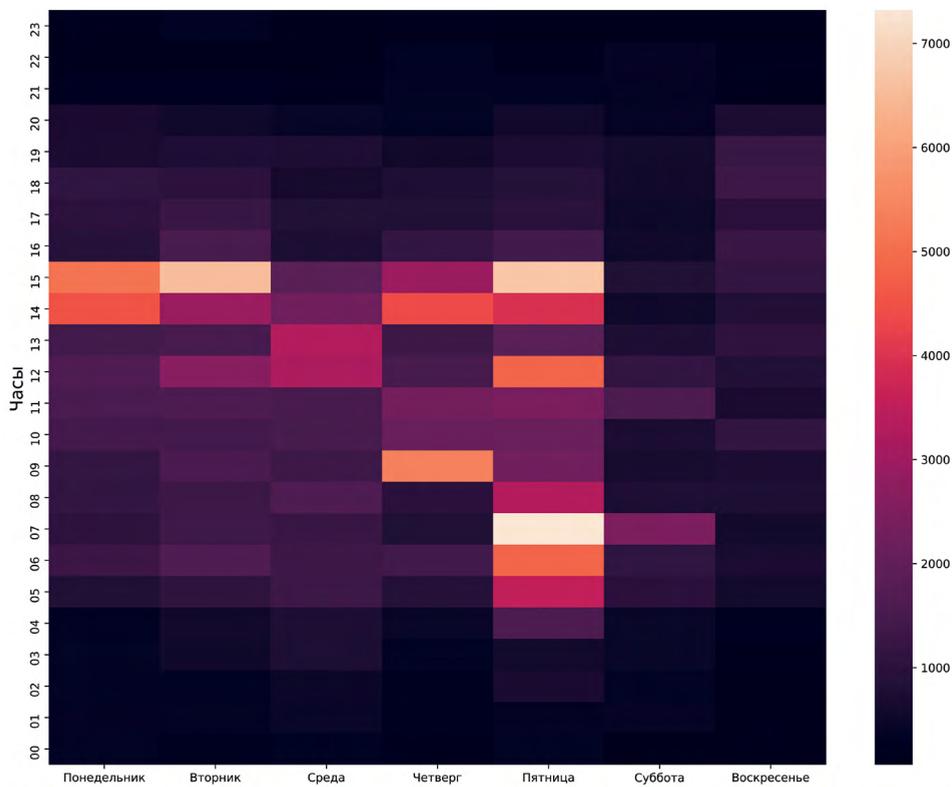


Рисунок 34. Визуализация распределения активности слушателей Центра подготовки РКЦТ (выполнение заданий) по дням недели и времени суток. Тепловая карта выполнена с помощью библиотеки Seaborn языка Python

4. Библиотеки статистического анализа, разработанные в разных языках программирования

R — язык, написанный специально для статистического анализа. 10–15 лет назад он был популярен в мире аналитики, но потом его потеснил Python. Изначально Python не разрабатывался для аналитики, но на практике оказался подходящим. Эти языки ценны разнообразием библиотек и пакетов — от простых, включающих традиционную статистику вроде сравнения средних значений и оценки корреляций, до машинного обучения. Кроме того, множество алгоритмов предназначено для обработки данных, включая очистку и трансформацию. Достаточно нескольких строчек кода, чтобы подключиться к источникам данных, подготовить их, рассчитать статистические оценки и построить визуализацию (см. рис. 34).

Одна из причин популярности аналитики на программных языках состоит в том, что пакеты анализа разрабатываются десятками тысяч людей на открытых платформах, быстро развиваются, содержат новые решения из мира аналитики и совершенно бесплатны. Создаваемые алгоритмы и решения легко внедрять (на сайт, в ИС, в веб-приложения).

5. Системы бизнес-аналитики (BI-системы) и дашборды

Понятие бизнес-аналитики (Business Intelligence, BI) возникло в 1970-е годы в финансовой сфере, где важно быстро принимать решения, анализируя большой объем данных. Для этого плохо подходили классические схемы проверки гипотез и статистики, зато отлично работала визуализация, что стало драйвером появления специализированных BI-систем. Последние 15 лет наблюдается демократизация таких систем: они стали доступны для среднего и малого бизнеса, а термин BI начал применяться в значении «анализ данных для бизнеса, для быстрых решений». Несмотря на свое название, BI-системы могут применяться не только в бизнес-среде, но и в сфере государственного управления.



По мнению основных аналитических агентств, к числу самых эффективных BI-платформ относятся Microsoft Power BI, Qlik и Tableau. Также существуют специфические системы, удобные для работы с государственной статистикой, например «Контур» (ее используют в «Росстате»).

BI-системы позволяют быстро собрать информацию, обработать ее, визуализировать и сделать выводы. Бизнес-аналитику можно представить как «завод» по получению из данных информации, значимой для принятия решений (рис. 35). BI превращает сырье (данные) в разнообразные



Рисунок 35. «Завод» по получению из данных информации

информационные продукты: «информацию», которая собирается и агрегируется в хранилищах данных (см. раздел 6.2); «знания», собираемые из запросов, сообщений и в результате применения аналитических инструментов; «планы», которые формируются «по кусочкам» из правил, закономерностей, моделей и схем, обнаруженных с помощью аналитических инструментов; и «действия», посредством которых бизнес-пользователи реализуют планы (которые генерируют события, в свою очередь дающие жизнь новому циклу).

У BI-систем масса плюсов, но есть и минусы, и один из них — затраты на разработку интерактивных отчетов для конечных пользователей. В BI-системах можно выделить три роли: разработчика, аналитика и пользователя.

Разработчик должен обладать широким набором компетенций и навыков, например знать язык запросов SQL и другие языки, разбираться в работе с данными. **Аналитик** выбирает нужные срезы, фильтрует данные и извлекает из них инсайты или готовит презентации, а затем показывает их лицам, принимающим решения. Руководитель должен поставить четкую задачу разработчику, объяснить свои потребности, а разработчик, в свою очередь, предложить возможные решения. **Конечный пользователь** (лицо, принимающее решение) не тратит время на фильтрацию данных. Он может сразу оценить ситуацию, взглянув на экран с дашбордом.

Это понятие (от англ. dashboard ‘приборная панель’) пришло в BI-аналитику из техники: первые дашборды действительно напоминали приборную панель самолета или автомобиля. Грамотно разработанный дашборд легко понимать, он не отвлекает внимание деталями и своевременно сигнализирует о проблемах.



«Разработка дашбордов и проверка качества данных, которые в них используются, отнимают много времени, но зато потом экономят много усилий — если дашборды дают информацию, действительно важную для принятия решений. В нашем Центре мы разработали несколько таких дашбордов: с данными о выполнении госзадания, которое выражается для нас в количестве обучающихся на программах; с кадровой статистикой (количество людей на данный момент в организации, сколько из них в отпуске, у кого день рождения) и с показателями, которые должны выполнять регионы. Когда все эти показатели собирают в ручном режиме, велика вероятность сбоя, задержки данных. Дашборд позволяет отслеживать ситуацию в режиме реального времени, видеть риски».

Ксения Ткачева, директор Центра подготовки РКЦТ

Выбор конкретного инструмента аналитики зависит от задач организации и от ее руководителей. У руководителя может возникнуть соблазн сразу построить сложную систему, внедрить в нее ИИ, но нужно сначала сделать шаг назад и понять, освоены ли базовые подходы к работе с аналитикой.



Предположим, организация занимается утилизацией мусора в разных районах области. Если задача — сравнить районы по количеству мусора и сопоставить эти данные с численностью населения, то вполне хватит Excel-таблицы с данными о каждой точке сбора твердых бытовых отходов и периодичности вывоза мусора. Но спрогнозировать с помощью этой таблицы, в каких точках потребность в вывозе мусора больше, а в каких — меньше, не получится, потребуется более серьезная система — SPSS или разработка на R или Python.

Коммерческие системы (комплексные универсальные статистические программы) покрывают большую часть потребностей в аналитике, в них удобно пользоваться оконным интерфейсом, труднее сделать ошибку, ниже «порог вхождения» (требования к опыту). Они распространены в научной среде и образовании, поэтому легче найти специалистов для работы с ними, но при этом они довольно дорогие. При использовании языков R или Python главным препятствием оказывается нехватка обученных сотрудников. Пакетов и библиотек много, они бесплатны, но подобрать нужные параметры, настроить — для этого требуется не просто знание языка, но и опыт подобной работы.

9.5.2 ПОДХОДЫ К ПОСТРОЕНИЮ СТРУКТУРЫ АНАЛИТИКИ

Задачи для аналитиков могут возникать как из объективных проблем и конкретных запросов (например, есть проблема текучки кадров), так и в более общем виде, как поиск точек роста и новых решений на основе данных. В этом случае важную роль играют РЦТ и CDO (см. раздел 5.1): они предлагают идеи улучшений за счет использования данных.



«Работа с данными — в большой степени творческая. Выдвинули гипотезу, подобрали данные, посмотрели алгоритмы — работают или нет. Дальше меняем гипотезу, данные или алгоритм. В этой работе нет и не может быть никакой методики, никакого стандарта. Пока у людей внутри не включится творчество, они реагируют на чужие идеи с подозрением: „пришли и хотят чего-то непонятного“. Но потом появляется человек с аналитическим складом ума, который не просто выполняет задачи, а начинает думать, что можно изменить».

Михаил Петров, директор департамента цифровой трансформации Счетной палаты РФ

Построение структуры аналитики в организации стоит начать со структурирования данных о потребностях в информации и аналитике в организации, с аудита покрытия этих потребностей текущими отчетами. Важно понять, кому из сотрудников какие данные нужны, почему и какой именно аналитики не хватает, выделить основные функциональные роли (должности, отделы, департаменты) и выяснить, какие решения они принимают, какая аналитика может помочь им в принятии решений (если надо — провести интервью). Нужно оценить важность и срочность аналитики (затраты на разработку серьезного исследования могут не окупиться) и понять, какие данные нужны, есть ли нужные данные, где они находятся.

Существует два основных подхода к построению структуры аналитики в организации: **централизованная аналитика и самообслуживание подразделений**.

Если организация выстроена жестко, иерархична, подразделения связаны между собой, если есть высокая чувствительность (сенситивность) данных (см. раздел 8.1), лучше идти в сторону централизованной аналитики. Если подразделения независимы друг от друга, ведут разные проекты в разной среде, то нет смысла объединять все в единую структуру. Эти подходы отчасти противоположны, но их комбинация возможна (более того, это оптимальный путь). При любом подходе крайне важно развивать культуру данных у сотрудников, мотивировать их к тому, чтобы тратить время на обучение и изучать инструменты аналитики.

ЦЕНТРАЛИЗОВАННАЯ АНАЛИТИКА

- › создается специальное подразделение, занятое производством отчетов;
- › от руководства и подразделений к нему поступают запросы;
- › сотрудники подразделения работают с разнообразными источниками и готовят витрины данных.

Плюсы:

- › высокая квалификация, сложный анализ;
- › хорошо для работы с чувствительными данными;
- › аналитики имеют доступ к корпоративному хранилищу, им легче получить доступ к локальным базам по всей организации;
- › аналитики могут обучать других пользователей, продвигать культуру данных.

Минусы:

- › неизбежен длинный бэклог: желающих больше, чем ресурсов;
- › хуже понимание потребностей подразделений;
- › унификация и, как следствие, пренебрежение некоторыми потребностями внутреннего заказчика;

- › дольше цикл разработки (понимание данных, уточнение задач, согласование), меньше оперативность;
- › до 90% аналитики может идти в корзину.

САМООБСЛУЖИВАНИЕ ПОДРАЗДЕЛЕНИЙ

- › в отделах или департаментах выделяются сотрудники, ответственные за аналитику, или вводятся должности аналитиков;
- › аналитика выполняется отдельно для каждого подразделения;
- › три типа пользователей: обычный пользователь (просто смотрит), продвинутый (опираясь на шаблоны, может делать анализ), аналитик.

Плюсы:

- › аналитики лучше понимают потребности, меньше дистанция от заказчика;
- › оперативно реагируют на запросы.

Минусы:

- › получается дороже (удельная стоимость отчета выше);
- › хуже качество аналитики;
- › стилевой разницей, у каждого свой тип визуализации;
- › нет «единой версии правды»: в каждом подразделении свои метрики, алгоритмы, их трудно сравнивать, нет общей картины;
- › аналитикам подразделений труднее дается профессиональный рост.

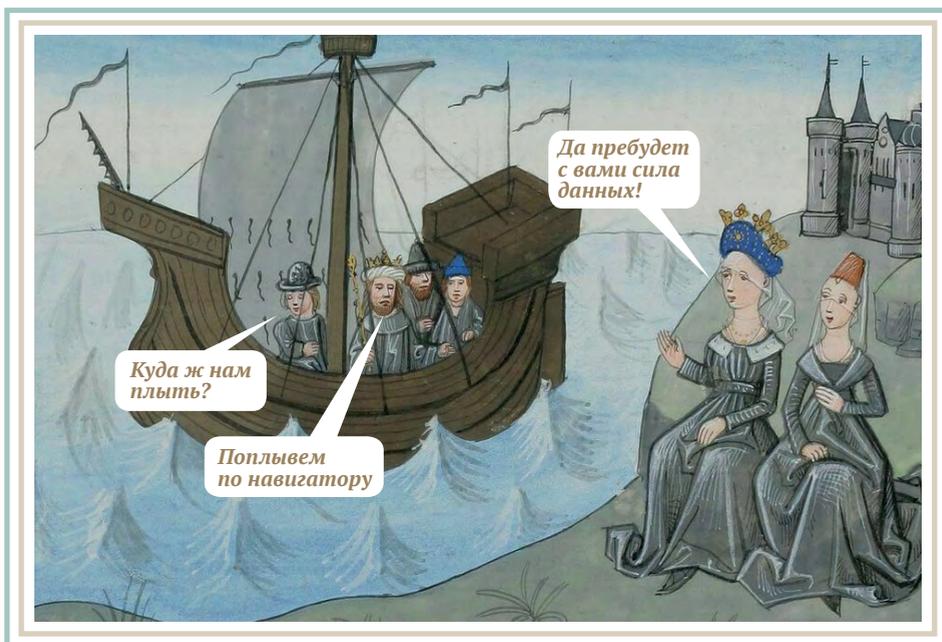
Важно выяснить, как именно отчет принесет пользу. Для этого полезно формулировать потребность в аналитике в виде «пользовательских историй». Самый простой план такой истории состоит из трех частей: а) кто, б) что именно система должна делать для пользователя, в) зачем.



а) Как руководитель б) я хочу видеть, сколько времени сотрудники моего департамента проводят за работой в среднем при работе в офисе и удаленно, в) чтобы лучше планировать гибридный формат работы и оперативнее реагировать на провалы мотивации отдельных сотрудников.

а) Как руководитель отдела транспортного планирования б) я хочу видеть динамику загрузки по основным маршрутам общественного транспорта, в) чтобы оптимально использовать резерв подвижного состава.

Использование аналитических инструментов в организации эффективно лишь при наличии правильно сформулированных задач, осознания необходимости описания и сбора данных, заботы об их качестве, — именно такой комплексный подход к управлению данными реализует фреймворк, описанный в приложении А.



ЗАКЛЮЧЕНИЕ. К ГОСУПРАВЛЕНИЮ НА ОСНОВЕ ДАННЫХ

 *Время чтения: 6 мин.*

Подводя итог трем основным блокам «навигатора для начинающих», авторы возвращаются к исходному тезису: у руководителей всех уровней, вплоть до самых верхних, появился ценный инструмент управления. Новое качество работы с данными позволяет увидеть, что сейчас приоритетно для организации, региона или ведомства, и принять решение на основе этого знания. Кроме описанных выше аспектов при работе с данными необходимо обратить внимание на еще два — этику и культуру работы с данными: они во многом будут определять успехи управления на основе данных в будущем.

Конечная цель освоения разнообразных методов работы с данными состоит в том, чтобы строить на основе данных госуправление всех уровней — от небольшого департамента до министерств и государства в целом¹⁹².

Раньше профессионализм руководителя определялся его разнообразным опытом, умением видеть тренды. Сейчас этого недостаточно, потому что есть данные и проверенные методы их анализа (подробнее см. главу 9).

¹⁹² По этой причине авторы не ограничатся настоящим навигатором и продолжат развивать тему в докладе об управлении на основе данных, выход которого планируется во второй половине 2022 года.



«Благодаря данным у государственных служащих разного уровня появляется возможность принимать управленческие решения, руководствуясь не только интуицией или опытом, который устаревает и вводит в заблуждение, а опираясь на актуальную и достоверную информацию. Если учесть, что государственные решения финансируются за счет налогов граждан, объективность и обоснованность этих решений приобретает для руководителей госсектора особое значение».

**Ксения Ткачева,
директор Центра подготовки РКЦТ**

В этом навигаторе авторы описали теоретические и практические стороны работы с данными, обсудили подходы к управлению ими и привели примеры использования данных в работе ведомств и организаций. Прodelав путь по всем девяти главам навигатора, читатель имел возможность увидеть, чем управление данными не является, и осознать главные **парадоксы данных**.

- ▶ Данные создаются внутри подразделений, но принадлежат организации в целом. Все данные описывают объекты, явления, события или другие данные во всей организации. Если смотреть на процессы организации как на сквозные, направленные на создание конечной ценности (товаров или услуг), становится очевидно, что данные не могут «принадлежать» отдельному подразделению. По этой причине в докладе мы избегали роли «владелец данных», заменяя ее на роль «ответственный за данные».
- ▶ Данные существуют в ИТ-системах, но управление данными — это не ИТ-функция. ИТ-функция создает ИТ-архитектуру (совокупность ИТ-систем и технологий организации), подобную градостроительному плану. Управление данными формирует архитектуру данных, которая задает связи потоков данных с объектами и процессами организации. Потоки данных движутся в ИТ-системах подобно транспортным потокам. Куда ехать, решают не строители, а водители, диспетчеры и ПДД.
- ▶ Организациям требуется не как можно больше данных, а быстрые и надежные «умные» данные. Большие данные сами по себе, без аналитики, не дают дополнительную ценность. Управление данными должно быть выстроено так, чтобы быстро предоставлять необходимый минимум информации для принятия решений, снижая уровень информационного шума.
- ▶ Управление данными — это не углубление процессной модели организации в попытке максимально подробно описать все процессы. Процессная модель — это план деятельности, который всегда хотя бы немного, но расходится с реальными процессами. Она определяет то,

что мы думаем о том, как должны работать процессы. Процессная модель напоминает карту города: карта никогда не поспевает за изменениями, которые в нем происходят. Модель данных — это описание того, что происходит в реальности. Она основана на потоках данных и исполнимых сценариях (исполнимый сценарий — пошаговый алгоритм действий сотрудников в ИС, оставляющий «цифровые следы» при его выполнении).

Авторы:



И. М. Добровольский



Е. Г. Потапова



Д. О. Теплякова

Завершая основную часть навигатора, отметим несколько тем, которые не вошли в его ограниченный объем, но при этом чрезвычайно важны для руководителя, работающего с данными. Это этика данных и культура работы с данными. Этика данных представляет собой формирующуюся отрасль прикладной этики, которая описывает ценностные суждения и подходы, применяемые при сборе, анализе и распространении данных¹⁹³, и тем самым обеспечивает использование данных надлежащим образом.

Примечательно, что в своде знаний DAMA-DMBOK раздел об этике предваряет все прочие разделы, что говорит об ее особой значимости. Для специалистов по работе с данными, помимо традиционно обсуждаемых проблем, существует еще и ряд узкопрофессиональных аспектов этой темы. Обычно в области этики данных выделяют три основных зоны рисков.

1. Защита персональных данных, за которую несут ответственность организации, собирающие эти данные (см. об этом главу 8).

2. Сохранение приватности, в частности при использовании метаданных, данных геолокации и других данных, которые содержат информацию о жизни человека. Особенно актуальной эта задача становится в случае объединения данных о человеке из разных источников. Этичный путь — это либо уничтожать данные сразу после того, как цель их сбора была достигнута, либо собирать минимальный объем данных¹⁹⁴.

3. Обработка датасетов для обучения ИИ, чтобы избежать дискриминации и принятия несправедливых решений. Для тех, кто анализирует данные и принимает решения на основе этого анализа, опасны любые формы подтасовки или злоупотребления статистикой, например:

- › манипуляции с хронологией — искажение результата из-за выборочного исключения и (или) включения временных точек ввода или регистрации данных;
- › неправильная визуализация — намеренное использование диаграмм и графиков, не подходящих для визуализации конкретных наборов данных, для формирования ложных представлений о данных (см. также раздел 9.4);

¹⁹³ Ср. британский фреймворк по этике данных. URL: <https://www.gov.uk/government/publications/data-ethics-framework>

¹⁹⁴ См. также об этом главу 4 «Приватность и защита персональных данных» в докладе «Этика и „цифра“: от проблем к решениям». М., 2021. URL: <https://ethics.cdto.center/2021/4-1-sposoby-obespecheniya-privatnosti>

- › предвзятость, систематические ошибки и искажения — подтасовка данных, выборочный поиск данных, подтверждающих конкретную гипотезу, необъективные выборки данных, влияние культурного контекста на объективность данных;
- › нехватка информации о первоисточнике данных, некачественные данные, ненадежные метаданные, отсутствие документированной истории изменений данных;
- › задержки в предоставлении данных.

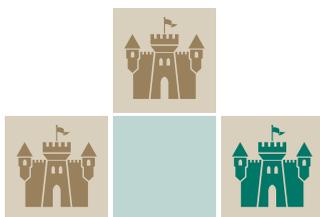
Все эти проблемы требуют внимания и усилий со стороны исследователей, экспертов и руководителей. Необходимо создавать стандарты и руководства по этичной работе с данными, разрабатывать механизмы надзора за соблюдением таких стандартов на практике — и на уровне организации, и на уровне государства в целом.



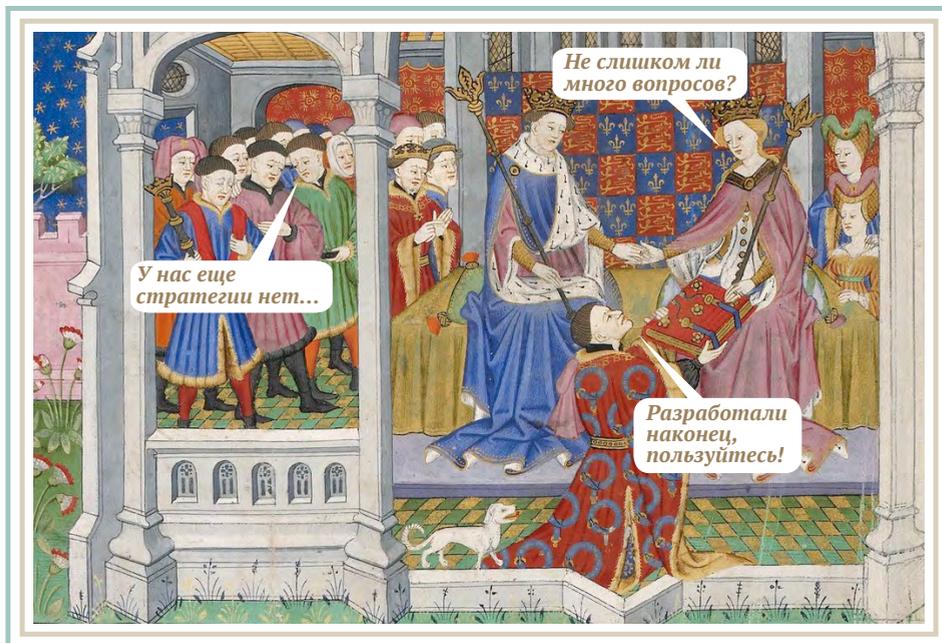
Центр подготовки РКЦТ уделяет большое внимание теме этики технологий. В 2021 году Центр собрал команду экспертов, которые создали фреймворк «Ответственная разработка цифровых решений», позволяющий организации провести диагностику этически проблемных зон создаваемых цифровых сервисов.

Вопросы культуры данных в организации не менее важны, чем тема этики. Это небыстрый и довольно сложный процесс для сотрудников, особенно в госсекторе. Взять в команду дата-сайентиста и завести озеро данных недостаточно. Предстоит планомерно вовлекать всех сотрудников в эту работу, терпеливо обучать, предлагать самые разные формы мотивации — от участия в хакатонах до премий и продвижения по службе. Необходимо помнить о том, что в первую очередь меняются люди, а не организации и технологии, поэтому у сотрудников должна быть возможность пройти все этапы процесса изменений: окончание предыдущего цикла и понимание того, какие практики следует сохранить, а с какими лучше расстаться; фазу перехода, в которой больше свободы и экспериментов, и новый старт — сами изменения. Перед внедрением новых систем, практик или методов нужно убедиться, что все участники хорошо знают свои роли, что регламенты и правила сформулированы ясно и не вызывают вопросов, а руководство дает четкие поручения и верит в необходимость перемен. Так процесс перехода будет плавным, а изменения — устойчивыми.

Приватность и защита ПДн, использование ИИ в обработке данных и в целом глобальный переход от управления данными в отдельных организациях, ведомствах и отраслях к государственному управлению на основе данных — темы, которые не уйдут из повестки в ближайшие месяцы и годы. Надеемся, наш навигатор приблизит читателей к пониманию роли данных в цифровой трансформации и будет способствовать наступлению этичного, удобного и человекоцентричного цифрового будущего.



ПРИЛОЖЕНИЯ



А. ФРЕЙМВОРК «РЕШЕНИЕ ЗАДАЧ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ»

Работа с задачей, при решении которой используются данные, — сложный и многоступенчатый процесс. Чтобы не упустить ни один из этапов и достичь поставленных целей, можно воспользоваться специальным инструментом — фреймворком. Авторы доклада, опираясь на опыт управления данными в организациях, предложили вариант такого фреймворка. Он подходит как для государственных структур, так и для коммерческих компаний.

 *Время чтения: 15 мин.*

А1 КАК ПОЛЬЗОВАТЬСЯ ФРЕЙМВОРКОМ

Фреймворк «Решение задач с использованием данных» предназначен для команд и организаций, решающих задачи с использованием данных. Основная цель фреймворка — структурировать работу с данными, проверить логику и последовательность действий, необходимых для успешного решения поставленной задачи.

При решении конкретной задачи, связанной с данными, есть опасность начать с более простых действий, например выбора информационных

систем, закупки оборудования. Однако прежде всего необходимо выяснить и описать фактические цели (ответить на вопрос «Зачем это делается?»), состояние организации, существующие в ней процессы, организационную структуру.

Фреймворк состоит из трех блоков: матрицы, опросника и смоделированного кейса с рекомендациями, который служит примером использования фреймворка. На основе матрицы составлен опросник из пяти этапов, который поможет командам, работающим с данными, проверить, обладают ли они достаточными ресурсами для решения задачи.

Работа с фреймворком начинается с постановки задачи. Можно считать, что задача поставлена хорошо, если она:

- › звучит максимально конкретно;
- › подразумевает использование данных конкретного типа;
- › относится к конкретной территории и предметной области.



Предположим, есть задача разобраться с объектами незавершенного строительства. Специалисты Счетной палаты сформулировали для анализа этих объектов следующие вопросы.

- › **В каком регионе самая сложная ситуация с незавершенным строительством?**
- › **В каком регионе максимальное количество аварийных жилых домов, подлежащих расселению?**
- › **Как получить сведения об участках, находящихся во владении граждан?**

Если необходимо получить многоступенчатый ответ, то вопрос сначала формулируют, а затем при необходимости уточняют:

- › **Как сравнить ситуацию с незавершенным строительством в Москве и Московской области?**
- › **Сколько объектов незавершенного строительства в области заморожено, а сколько продолжает строиться за счет средств местных бюджетов?**
- › **Как сравнить количество замороженных объектов и обманутых дольщиков для объектов малоэтажного и многоэтажного строительства в Московской области?**

После того как задача поставлена, можно приступать к прохождению опросника. Он состоит из двадцати вопросов, на которые можно ответить «да» или «нет». Чем больше ответов «да», тем с большей вероятностью задача будет решена успешно.

Если для вашей задачи неактуально, например, использование того или иного инструмента или присутствие той или иной роли в команде, отвечайте «неактуально». Если вы ответили «нет» на один

Авторы:



О. М. Гиацинтов



И. М. Добровольский



А. А. Малахов



В. А. Сазонов



Д. О. Теплякова

из вопросов, это значит, что в процессе решения вашей задачи есть пробел, который необходимо ликвидировать, прежде чем двигаться дальше.

Использовать опросник можно как при работе с данными внутри организации, так и в случае привлечения подрядчиков. Желательно проводить обсуждение в группе под руководством CDO с привлечением сотрудников подразделений, работающих с данными, и сотрудников функционального подразделения, которому необходимо решить задачу.

A2 ИНСТРУМЕНТЫ ФРЕЙМВОРКА

A2.1 МАТРИЦА

Матрица (табл. 7) представляет собой визуальный инструмент, где собраны элементы работы с любой задачей, для решения которой используются данные: этапы решения задачи (постановка задачи — описание данных — сбор данных — использование данных — результат) и три группы параметров, которые присутствуют на каждом этапе (документы, роли и инструменты).

Для удобства в матрице размещены ссылки на соответствующие разделы навигатора — если хотите больше узнать о том или ином компоненте процесса, вы можете обратиться к главам, в которых они описаны.

Таблица 7. Матрица решения задач с использованием данных



	Задача	Описание данных	Сбор данных	Использование данных	Результат
Документы (см. разделы 1.1, 3.1, 4.4, 7.3)	Требования	Методология руководства данными, методики проверки качества, регламенты работы с критичными данными	Методология и регламенты построения хранилища и озера данных, управления качеством данных, управления НСИ	Методики и регламенты формирования отчетности и проведения аналитических работ	Готовые отчеты и результаты анализа
Роли (см. разделы 5.1, 6.1, 8.3)	Пользователь данных, эксперт	CDO, архитектор данных, стюард данных, аналитик данных	СIO, архитектор данных, системный аналитик, разработчик, тестировщик	Исследователь данных (Data Scientist), аналитик	Пользователь данных
Инструменты (см. разделы 6.2, 7.2, 8.3, 9.5.1)	Супермаркет данных	Бизнес-гlossарий, каталог данных, решения DQ, категоризация критичных данных	Хранилище и озеро данных, Data Mesh Architecture, ETL- и DQ-инструменты, SQL, управление мастер-данными (MDM), СУБД, Hadoop	Бизнес-аналитика, интеллектуальный анализ данных (Data Mining), решения Self-service Data Preparation	Супермаркет данных

А2.2 ОПРОСНИК

Этап I. Постановка задачи

1. Разработана ли у вас стратегия ЦТ и соответствует ли она стратегии развития организации в целом?
2. Разработан ли в вашей организации подход к выбору задач и (или) построению кейсов?
3. Есть ли у вас КПЭ по работе с данными и соответствуют ли они КПЭ организации в целом?

Этап II. Описание данных

1. Есть ли в вашей организации стратегия управления данными?
2. Есть ли в вашей организации руководитель по работе с данными (CDO) / департамент управления данными (дата-офис)?
3. Есть ли в вашей организации регламент управления данными?
4. Выбрали ли вы инструменты для:
 - а) ведения концептуальных и логических моделей данных (глоссария);
 - б) инвентаризации физических моделей данных и процессов обработки данных;
 - в) проверки качества данных;
 - г) категоризации критичных данных;
 - д) приема запросов на данные?
5. Сформирован ли подход к выбору ответственных за данные?
6. Назначены ли:
 - а) эксперты по данным в своих функциональных подразделениях (SME);
 - б) комитет по данным?

Этап III. Сбор данных

1. Есть ли у вас положения и (или) регламенты, описывающие:
 - а) формирование хранилища данных и (или) витрины данных;
 - б) формирование озера данных;
 - в) управление нормативно-справочной информацией (НСИ);
 - г) обеспечение качества данных;
 - д) интеграцию данных?
2. Выбрали ли вы инструменты для:
 - а) формирования хранилища данных и (или) витрины данных;
 - б) формирования озера данных;
 - в) управления НСИ;
 - г) обеспечения качества данных;

- д) интеграции данных;
 - е) создания физической модели данных?
3. Разработали ли вы политику информационной безопасности (ИБ) в отношении персональных данных, коммерческой, государственной и иных видов тайн, использования тестовых сред и т. п.?
 4. Разработали ли вы подход к реализации программных проектов?
 5. Выбрали ли вы инструменты для:
 - а) мониторинга в части ИБ;
 - б) создания тестовых сред;
 - в) управления ролевой моделью и доступом?

Этап IV. Использование данных

1. Подготовлены ли инструкции по созданию и использованию отчетности?
2. Есть ли регламент использования данных?
3. Выбрали ли вы инструмент BI для:
 - а) проведения аналитических работ;
 - б) подготовки данных;
 - в) визуализации данных?
4. При необходимости проводить аналитические работы: есть ли в команде исследователь данных?

Этап V. Результат

1. Провели ли вы анализ результатов и подходов?
2. Продумали ли вы тиражирование на другие подразделения и (или) масштабирование?

A3 ПРИМЕР ПРИМЕНЕНИЯ ФРЕЙМВОРКА: УПРАВЛЕНИЕ ТУРПОТОКАМИ

Представьте, что вы — руководитель отдела в региональном ведомстве, которое занимается развитием туризма. Перед вашим отделом поставлена задача выяснить, сколько туристов в реальности посещает регион, и понять детальную структуру турпотока, чтобы подготовить основу для принятия решений о строительстве инфраструктурных объектов.

Всех данных, необходимых для решения задачи, у вас в организации сейчас нет. Вы собираетесь обратиться к подрядчикам — в аналитическую компанию. В этот момент следует взять паузу и провести аудит, ответив на вопросы нашего фреймворка. Какую картину вы с большой вероятностью обнаружите?

На этапе I (постановка задачи) положительный ответ, скорее всего, будет дан лишь на вопрос о наличии стратегии ЦТ. На этапе II (описание данных) положительных ответов не будет. На этапе III большинство ответов будет «нет»; у вас есть только управление НСИ и политика ИБ. Что касается этапа IV (использование данных), у вас, возможно, имеется один из VI-инструментов и регламент постановки задач, связанных с данными, и получения результатов. На этапе V вы, вероятно, проводите анализ подходов и результатов, но вряд ли масштабируете полученные решения, если в этом не было прямой необходимости. Как показывает опыт авторов доклада, такая картина характерна для значительной части РОИВ.

Что можно сделать на каждом этапе, чтобы задача, стоящая перед ведомством, была успешно решена в положенный срок?

На этапе I (постановка задачи)

- ▶ Если стратегии ЦТ нет, ее можно разработать, используя опыт администрации области, региона или ФОИВ и опираясь на стратегии ЦТ субъектов, а также на аналитический доклад «Стратегия цифровой трансформации: написать, чтобы выполнить»¹⁹⁵.
- ▶ В стратегии ЦТ (или в приложении к ней) должен быть описан процесс постановки задач и обработки поступающих кейсов. Нужно понимать, что оценка турпотоков может стать первым этапом на пути развития региона в целом: дальше нужно будет решать, где строить магазины, кафе и автозаправки. Возможно, появятся новые рабочие места, нужно будет наращивать инфраструктуру. Соответственно, выбор задач нужно соотнести со стратегией.

На этапе II (описание данных)

- ▶ Необходимо разработать стратегию управления данными и КПЭ в соответствии со стратегией ЦТ организации.
- ▶ На начальном этапе должен быть определен сотрудник, ответственный за ЦТ; скорее всего, он будет отвечать и за данные. При большом объеме работы через некоторое время имеет смысл создать отдельную должность CDO (см. раздел 5.1).
- ▶ Регламент управления данными формируется из стратегии. В нем должно быть отражено, какие данные описывать, как описывать, как обращаться за данными к поставщикам.
- ▶ Если данные напрямую переходят в аналитическую компанию, то в ведомстве не нужно создавать специальную инфраструктуру. Достаточно разработать регламент контроля периодичности поступления данных в аналитическую компанию и отслеживать этот процесс. Если же часть данных будет находиться в ведомстве, нужно, чтобы в нем постепенно появлялись инструменты управления данными. В любом случае необходимо иметь систему, в которой будет описано все, что ведомство может запросить у аналитической

¹⁹⁵ Стратегия цифровой трансформации: написать, чтобы выполнить / под ред. Е. Г. Потаповой, П. М. Потева, М. С. Шклярчук. М.: РАНХиГС, 2021. URL: <https://strategy.cdto.ranepa.ru>

компании и предоставить ей. В организации должен быть налажен процесс приема запросов на данные.

- › Необходимо сформировать подход к выбору ответственных за данные. За это не должен отвечать подрядчик, ведомство само обязано следить за тем, чтобы в аналитическую компанию передавались качественные данные и чтобы это происходило регулярно, с нужной периодичностью.
- › Должны быть назначены эксперты по данным внутри организации, но можно привлечь и сторонних консультантов. Создавать комитет по данным имеет смысл только тогда, когда работы станет много и возникнет такая потребность.

На этапе III (сбор данных)

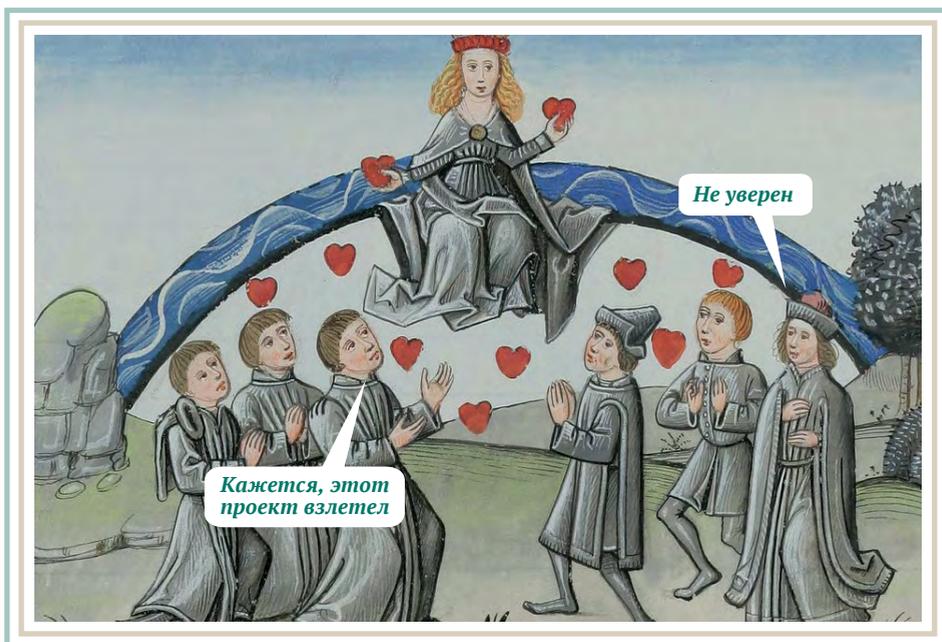
- › В ведомстве должно быть положение о хранилище первичных, собственных данных, положение по управлению НСИ, качеству данных и интеграции данных, поскольку необходимо регламентировать процессы передачи и приема данных при взаимодействии с аналитической компанией. Регламенты для озера данных достаточно иметь аналитической компании — подрядчику.
- › Инструменты для хранилища данных, обеспечения качества данных, интеграции данных и создания физической модели данных нужны только в том случае, если данные хранятся в самом ведомстве. Для управления НСИ лучше использовать промышленное решение.
- › Даже если программные проекты выполняет подрядчик, организации все равно нужно разработать подход к их выполнению.
- › Необходимо выбрать инструменты для мониторинга в части ИБ (см. раздел 8.3) и управления ролевой моделью и доступом; аналитическая компания должна ориентироваться на правила, которые установит ведомство.

На этапе IV (использование данных)

- › Необходимо разработать инструкции по созданию и использованию отчетности, а также регламент использования данных.
- › Возможно, выбранный инструмент BI (см. раздел 9.5.1) нужно будет адаптировать. Инструменты для аналитических работ и подготовки данных могут быть у аналитической компании, как и сотрудник, выполняющий роль исследователя данных.

На этапе V (результат)

- › Непрерывно анализировать результаты и корректировать подходы в зависимости от этих результатов.
- › Для описанного выше кейса необходимо масштабирование на другие подразделения. Нужно продумать подход к такому масштабированию.



Б. ОПЫТ УПРАВЛЕНИЯ ДАННЫМИ В РОССИИ

 *Время чтения: 40 мин.*

Осваивая современные методы управления данными, руководители и команды цифровой трансформации получают инструменты для поддержки управленческих решений, надежную аналитику, оптимизируют процессы в своих организациях. Собранные в этом приложении успешные кейсы семи российских ведомств и регионов показывают, что налаженная работа с данными может изменить органы государственного управления и выступить драйвером изменений в социальной сфере.

Б1 Анализ эффективности единой государственной системы отчетности в Республике Татарстан

Название проекта Анализ эффективности единой государственной системы отчетности «Отчеты ведомств» информационного портала «Открытый Татарстан»

Год 2020

Название организации Центр цифровой трансформации Республики Татарстан (далее — ЦЦТ РТ); комитет по социально-экономическому

Авторы раздела:



Р. И. Абдрахманов



А. А. Орлова



Д. О. Теплякова

мониторингу Республики Татарстан (далее — КСЭМ РТ); центр информационных технологий Республики Татарстан (далее — ЦИТ РТ)

Цель проекта Увеличить эффективность при составлении отчетов, уменьшить трудозатраты.

Команда

- › Исполнительный директор ЦЦТ РТ
- › Начальник отдела управления данными ЦЦТ РТ
- › Руководитель проектов ЦЦТ РТ
- › Аналитик данных ЦЦТ РТ
- › Департамент информационной безопасности ЦИТ РТ
- › Начальник отдела информационных ресурсов и технологий КСЭМ РТ
- › Ведущий специалист отдела информационных ресурсов и технологий КСЭМ РТ

К началу проекта команда была сформирована. Основной ее состав (роли 1–5) участвует в большинстве прикладных проектов отдела управления данными ЦЦТ РТ.

Результат По итогам проекта был выработан ряд рекомендаций, в частности: автоматизировать заполнение показателей, используемых в наиболее популярных отчетах; переводить непопулярные отчеты в архив, чтобы избежать затрат на их сопровождение.

Интересный факт В единой государственной системе отчетности «Отчеты ведомств» информационного портала «Открытый Татарстан» (далее — ЕГСО ОВ) накапливается и обрабатывается более 4000 показателей по различным направлениям деятельности Республики Татарстан, таким как образование, здравоохранение, экономика, промышленность, строительство, транспорт и дороги, общественная безопасность, статистика, финансы и бюджет, экология, потребительский рынок и др. Показатели используются при построении отчетных форм, которых в ЕГСО ОВ около 500. Более 1300 сотрудников органов исполнительной власти и местного самоуправления занимаются актуализацией, проверкой и публикацией отчетов.

Предпосылки

ЕГСО ОВ информационного портала «Открытый Татарстан» была создана в 2013 году¹⁹⁶. Проект анализа ее эффективности родился из гипотезы о том, что не все отчеты и показатели в системе одинаково эффективны и полезны. Есть показатели, которые используются в востребованных отчетах с большим количеством просмотров. А у некоторых отчетов просмотров так мало, что затраты на сопровождение этих отчетов не окупаются.

¹⁹⁶ Постановление кабинета министров Республики Татарстан от 27.08.2013 № 603 «О создании единой государственной системы отчетности „Отчеты ведомств“ информационного портала „Открытый Татарстан“» // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/917052336>



Рисунок 36. Пример главной страницы дашборда

Особенности проекта

Для анализа ситуации на основе данных было выбрано BI-решение Qlik Sense Enterprise. Решение имеет широкие функциональные возможности, как в части ETL (см. раздел 6.2), так и в части работы с моделью данных и визуализацией (см. рис. 36). Это решение можно было развернуть в том же центре обработки данных (далее — ЦОД), где развернута ЕГСО ОВ.

Команда проанализировала ряд метрик: трудозатраты сотрудников на отчет в часах, общее количество просмотров отчета с учетом его периодичности, еще некоторые другие (вспомогательные) метрики. Отношение количества просмотров к величине трудозатрат было названо **индексом эффективности отчета**. Отчеты с индексом ниже среднего были отнесены к красной зоне, остальные — к зеленой. Отчеты красной зоны были проанализированы более детально; среди отчетов со сравнительно высокими показателями просмотров были выявлены наиболее трудозатратные.

Безопасность

Qlik Sense был развернут в контролируемой зоне ЦОД — той же, где развернута ЕГСО ОВ, — что позволило сэкономить на приобретении средств криптографической защиты информации. Также затраты на информационную безопасность были снижены благодаря исключению из обработки персональных данных пользователей.

Подключение пользователей к BI было реализовано через виртуальную частную сеть, а подключение к веб-серверу Qlik Sense защищено шифрованием SSL. Дополнительно были настроены межсетевые экраны, препятствующие несанкционированному доступу на уровне локально-вычислительной сети ЦОД.



М. В. Петров

Б2 Система управления данными в Счетной палате РФ

Название проекта Внедрение системы управления данными

Год 2021

Название организации Счетная палата Российской Федерации

Цель проекта Существенно повысить качество данных и удобство их использования, автоматизировать их обработку, изменить культуру работы с данными в организации.

Команда Департамент цифровой трансформации Счетной палаты осуществляет общее руководство проектом; институт специалистов по описанию данных (дата-стюардов) в составе службы управления данными формирует ландшафт данных Счетной палаты.

- Результаты**
1. Создана система управления данными как модуль цифровой платформы Счетной палаты. В нем ведется описание данных, а также настраиваются правила, которым должны соответствовать данные.
 2. Создана служба управления данными; ей поручено обеспечивать инспекторский состав качественными данными. В состав службы входят дата-стюарды (см. о них разделы 4.4 и 5.1), дата-инженеры.

Интересный факт Дата-стюарды составляют карты данных, описывая по специальной методике ИС — источники данных, готовят описания наборов данных и витрин данных (см. рис. 37). Чтобы создать и успешно интегрировать правила управления качеством в систему, дата-стюард должен быть экспертом в предметной области и уметь преобразовывать описание правил в алгоритмы. Рассматривались два варианта создания службы — централизованный и децентрализованный, при котором дата-стюарды рассредоточены по отдельным департаментам. Был выбран второй вариант.

О проекте

В Счетной палате РФ создана цифровая платформа, объединяющая все основные бизнес-процессы организации. Ее составная часть и основа — система управления данными (СУД).

Предпосылки

Ранее данные накапливались в том виде, в котором они поступали в Счетную палату (в том числе разноформатные, неполные, некорректные, устаревшие). На проверку и очистку уходило много времени, приходилось дорабатывать системы сбора данных. Нужно было создать принципиально новую систему управления данными, которая позволяла бы описывать весь ландшафт данных, доступных для работы, быстро подключать новые источники, настраивать правила проверки качества данных.

Технология

Цифровая платформа в целом имеет трехслойную архитектуру.

- › Верхний слой непосредственно связан с бизнес-процессами: он обеспечивает проведение проверок, работу инспекторов, взаимодействие с экспертами.
- › Средний слой — озеро данных, хранящее данные, которые поступают из внутренних и внешних источников.
- › Глубинный слой — источники, из которых данные попадают в озеро данных.

Система управления данными Счетной палаты РФ (рис. 38) — основной элемент цифровой платформы Счетной палаты, позволяющий обеспечить централизованный учет, инвентаризацию информационных ресурсов, а также нормализацию, обогащение и качество данных.

К ключевым процессам в СУД можно отнести:

- › формирование карт данных;
- › описание подключения к источникам данных;
- › описание статуса первичной загрузки наборов данных;
- › мониторинг загрузки и обновления наборов данных;
- › создание правил качества данных и контроль качества данных;
- › мониторинг результатов контроля качества данных.

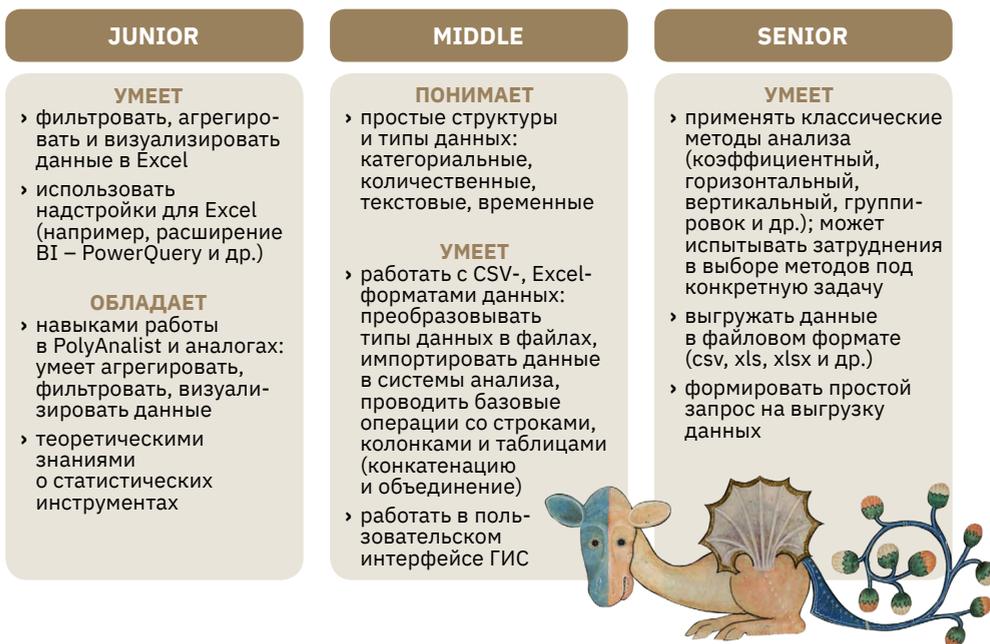


Рисунок 37. Модели компетенций дата-стюардов, разработанные в Счетной палате РФ

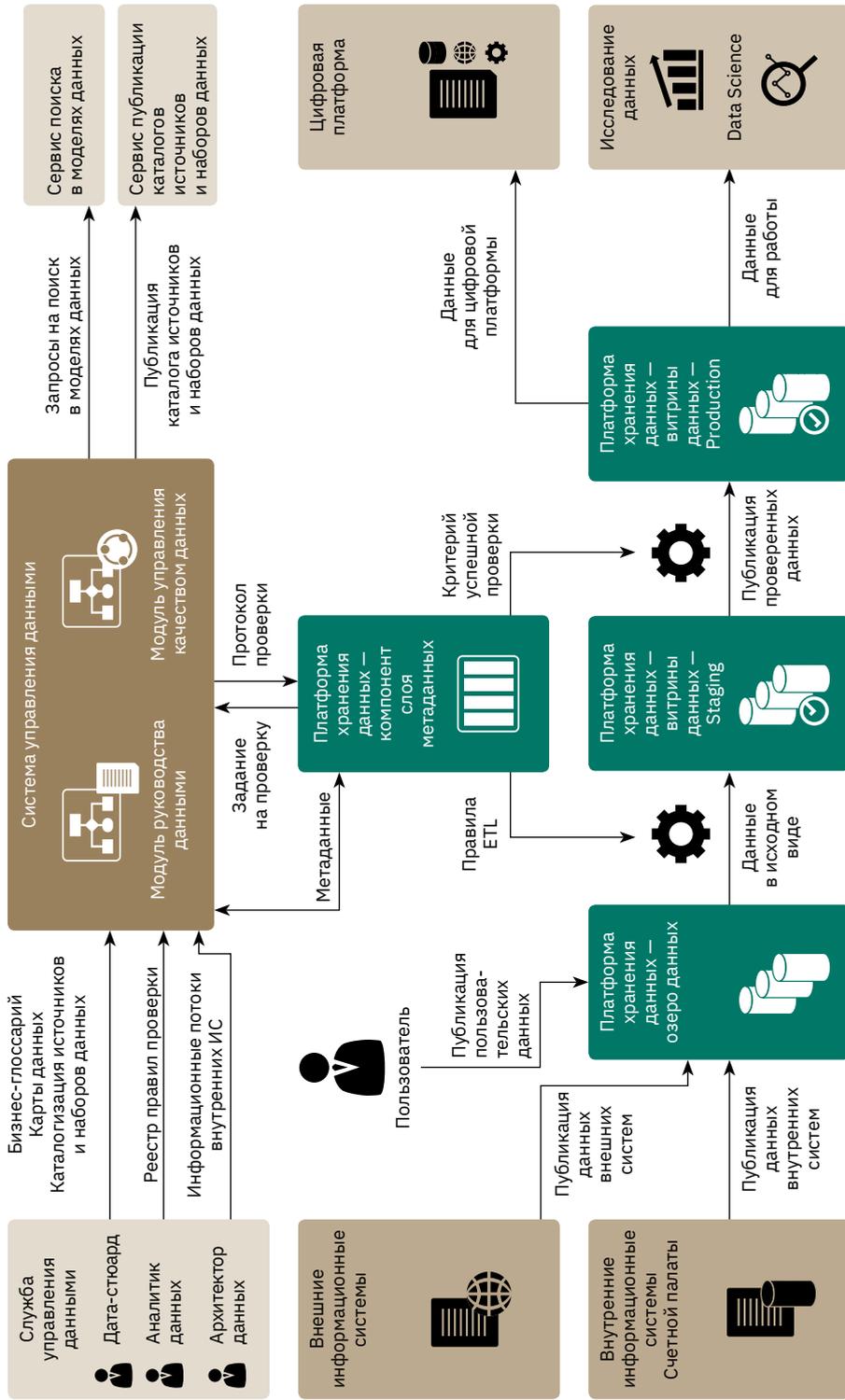


Рисунок 38. Система управления данными Счетной палаты Российской Федерации

Все процессы реализованы в едином удобном интерфейсе, что позволяет быстрее находить уже загруженные и обработанные данные, понимать, какие источники данных имеются в распоряжении, по каким ведется работа.

Особенности проекта

В СУД используются карты данных и инструменты проверки качества данных, контроль качества данных стандартизирован. Особое внимание уделяется таким метрикам данных, как своевременность и волатильность, точность, валидность, согласованность (см. раздел 7.3). Взаимодействие подразделений является в СУД сквозным процессом: нет необходимости направлять служебные записки, письма между подразделениями.

Система постоянно действующей проверки качества эффективна при большом и относительно неизменном потоке данных. Если данные имеют разный формат, разные цели, поступают спорадически, система обработки данных вряд ли будет эффективна. В этом случае проще проверять качество данных вручную.



Прежде чем загрузить в озеро данные загранпаспорта, их проверяют на валидность (дата окончания действия паспорта должна быть в будущем, а не в прошлом) и корректность (разность между датой окончания срока действия паспорта и датой его выдачи должна быть ровно 10 лет). Если эти условия не соблюдаются, значит, у паспорта закончился срок действия, он недействителен или имеются ошибки ввода данных.

При обнаружении ошибки дата-стюард во взаимодействии с инспекторским составом принимает решение повторно запросить данные у источника или, если позволяет ситуация, оперативно исправить данные. В СУД создаются инструменты, поддерживающие такие действия.

Результаты

Определены источники знания и инструментарий для работы с данными. Созданы отдельные профили компетенций для дата-стюардов (см. раздел 5.1) и остальных сотрудников, работающих с данными. Чтобы ускорить работу и тратить меньше ресурсов, выстраивается взаимодействие с Минцифры России в части гармонизации методологических подходов к управлению данными, в целях объединения описания данных в НСУД (см. раздел 3.1.2) с описаниями в СУД Счетной палаты РФ и создания единого информационного поля. СУД находится в промышленной эксплуатации, проведена работа по формированию структуры и укомплектованию службы управления данными, обучены ключевые сотрудники департаментов и инспекций Счетной палаты РФ. К концу 2021 года СУД оперировала 908 источниками данных, 634 наборами данных и 80 картами данных.

Авторы раздела:



И. А. Анисимова



А. В. Архипов



А. А. Орлова

Б3 Использование результатов хакатона в практике Роструда

Название проекта Хакатон PandemicDataHack

Год 2020

Название организации Федеральная служба по труду и занятости

Цель проекта Создание моделей и проверка гипотез для развития сервисов единой цифровой платформы (ЕЦП) «Работа в России»; исследования в сфере рынка труда в период пандемии.

Команда Руководитель и заместитель руководителя проекта, постановщики задач, аналитики данных, дата-инженеры, координаторы треков, отраслевые эксперты, медиакоманда, специалисты технической поддержки.

Результат Созданы прототипы и модели сервисов ЕЦП «Работа в России» (для служб занятости, граждан и работодателей). Результаты исследований используются для трудоустройства безработных граждан.

Интересный факт Более 2500 центров занятости используют в своей деятельности сервисы ЕЦП.

О проекте

В декабре 2020 года Роструд совместно с АНО «Центр перспективных управленческих решений» (ЦПУР) провел онлайн-хакатон по решению задач в сфере труда и занятости во время коронавируса PandemicDataHack¹⁹⁷. Его участникам были предложены четыре направления работы (трека). Участники исследовательского трека на основе административных данных портала «Работа в России» должны были определить:

- › как изменился профиль безработного в службе занятости;
- › какие группы населения больше всего рискуют потерять работу в кризис;
- › в каких отраслях рынки труда более устойчивы в кризис;
- › какие факторы влияют на трудоустройство граждан во время пандемии.

Участники трека машинного обучения на основе массивов вакансий и резюме создавали модели определения релевантной зарплаты для соискателя. Роструд планировал использовать результаты хакатона для создания новых сервисов в ЕЦП «Работа в России», а также для процессов модернизации служб занятости (проект «Служба занятости 2.0»).

К данным, подготовленным для хакатона, были применены процедуры снижения детализации и анонимизации, чтобы исключить риск раскрытия чувствительных данных, но не потерять их ценность для исследований.

¹⁹⁷ См. официальный сайт хакатона PandemicDataHack. URL: <https://pandemicdatahack.ru>

Предпосылки

В последние годы в службах занятости многих стран, в том числе в России, появился запрос на кастомизацию услуг — формирование предложений, которые в наибольшей степени подходят конкретному гражданину, той или иной категории (например, молодежи или людям с инвалидностью). Для развития сервисов специалистам Роструда важно анализировать не только общие тренды, но и более частные закономерности: по отдельным группам граждан, территориям, социальному статусу, возрасту и т. д. Потребность в таких подходах еще более усилилась с 2020 года, когда уровень регистрируемой безработицы впервые в российской истории существенно приблизился к уровню общей безработицы, рассчитываемой по методологии Международной организации труда.

Весной 2020 года на портале «Работа в России» появилась возможность онлайн-обращения за услугами. В результате за содействием в поиске работы обратились несколько миллионов граждан, объем данных значительно возрос. На рынке труда и занятости произошли существенные перемены, ситуация потребовала оперативных решений, создания новых сервисов для граждан на ресурсах Роструда, выработки новых подходов, основанных на анализе данных. Для создания моделей будущих сервисов и проверки гипотез в декабре 2020 года был инициирован хакатон с привлечением молодых исследовательских команд.

Особенности

По данным о вакансиях и резюме участники трека машинного обучения строили модели, предсказывающие уровень зарплаты, на которую может рассчитывать соискатель. Хакатон проводился на платформе Kaggle.



Kaggle (kaggle.com) — система организации конкурсов и социальная сеть специалистов по обработке данных и машинному обучению. На этой платформе исследователи разных уровней могут опробовать свои модели анализа данных на реальных задачах. Ресурс дает возможность получить профессиональный опыт, авторов лучших моделей награждают денежными призами.

Для более точного соответствия между вакансиями и резюме исследователи и аналитики данных предложили включить в число собираемых данных сведения о поведении пользователей на портале: какие вакансии просматривает автор резюме, как меняется интенсивность его поиска. На портале появился блок, позволяющий собирать такие сведения.

Результаты

1. Алгоритмы, опробованные на хакатоне, стали отправной точкой внедрения технологий машинного обучения для работы с вакансиями и резюме. В 2021 году в ЕЦП «Работа в России» был запущен инновационный сервис

CV-matching, позволяющий центрам занятости оперативно подбирать как максимально релевантные вакансии для соискателей, так и резюме кандидатов на открытые позиции в компании. Этим сервисом, созданным с использованием компонентов ИИ, уже могут пользоваться граждане и работодатели.

2. Исследования участников хакатона использовались в работе центров занятости с различными категориями граждан, в том числе с молодежью.

3. Проведение хакатона на детальных данных, ранее недоступных широкому сообществу исследователей и дата-аналитиков, позволило:

- › по-новому взглянуть на услуги и сервисы Роструда, собрать гипотезы и предложения, которые легли в основу концепции по совершенствованию действующих сервисов;
- › активно внедрять методы доказательной политики на рынке труда;
- › строить сообщество пользователей, анализирующих данные ведомства, которых в дальнейшем можно привлекать в качестве внешних экспертов.

Авторы раздела:



Д. А. Суржко



М. В. Туманова



Ю. А. Шамаев

Б4 Геоаналитическая платформа ВТБ

Название проекта Геоаналитическая платформа

Год 2020–2021

Название организации Банк ВТБ

Цель проекта Создание инструмента поддержки принятия решений

Команда

- › Инфраструктурная группа разрабатывает интерфейс, осуществляет поддержку серверов, ведет договорную базу.
- › Команда исследователей данных разрабатывает дизайн продукта и модели ИИ.
- › Платформа больших данных (совместное предприятие ВТБ и «Ростелекома») взаимодействует с заказчиками.

Результат Геоплатформа представляет собой универсальный геоаналитический инструмент, способный решать широкий спектр задач: от размещения точек продаж до управления рисками.

Интересный факт Геоплатформа стала базой для принципиально нового продукта на рынке — сервисов Geo-embedding, основанных на применении нейронных сетей. Они позволяют DataScience-команде платформы объединять усилия с DS-командами клиентов, вместо того чтобы конкурировать с ними или заменять их.

О проекте

В банке ВТБ создана геоаналитическая платформа, которая объединяет как собственные данные банка, полученные в результате геотегирования использования банковской карты, так и данные других организаций:

Таблица 8. Варианты сотрудничества в сфере обработки данных

Вариант сотрудничества	Недостатки
Компетенция по обработке данных полностью передается в ВТБ	Неудобно для партнеров, так как они должны передать все свои данные и компетенции в критически важной области (например, полностью отдавая на аутсорс вопрос размещения точек продаж)
Сырые данные и функционал геоплатформы полностью передаются партнерам	Дорого для партнеров, так как исходные данные для платформы стоят дорого. Возникает дилемма: развить в компании компетенцию по геоаналитике или полностью полагаться в этом на партнеров
ВТБ и партнеры строят свои независимые модели, а потом налаживают их взаимодействие	Для получения эффективного результата данные должны взаимодействовать между собой нелинейно

сотовых операторов, интернет-провайдеров, портала сделок с недвижимостью. Геоплатформа — инструмент, способный решать широкий спектр задач: от прогнозирования спроса, размещения точек продаж, рекламы и оптимизации клиентопотока до управления рисками. Решение может применяться и в сфере госуправления, например для аналитики туристических потоков, оценки инвестиционной привлекательности, а также для решения локальных задач: выбора места для нового отделения МФЦ или оптимизации его загрузки, прогнозирования востребованности отдельных госуслуг.

По сравнению с экспертной оценкой и моделью на основе открытых данных платформа обладает большей точностью. Она вызвала интерес сторонних организаций; оставалось выбрать способ сотрудничества (см. табл. 8). Требовалось найти решение, позволяющее достичь синергии¹⁹⁸ на уровне данных и при этом сохранить комфортное распределение компетенций и первичной информации между партнерами. Подход Data Fusion (сквозное объединение данных) стал очередным шагом в предиктивной аналитике, который помогает бизнесу создавать персонализированные предложения и прогнозировать потребности клиентов. В 2021 году в ВТБ этот подход был реализован в технологии Geo-embedding (геоэмбеддинга, или векторного представления клиентов), позволяющей решать проблемы, которые возникают при объединении данных.

Особенности платформы

- › Уникальный набор данных, которые удалось собрать в банке и с помощью партнеров. Более 200 первичных геослоев и около 1000 производных позволяют делать прогнозы на качественно новом уровне детализации и точности. На геоплатформе можно решать задачи эффективного размещения точек продаж, рекламы, оптимизировать поток клиентов.

¹⁹⁸ Модели на основе машинного обучения невозможно построить независимо, объединив только прогнозы; максимальная эффективность достигается при объединении данных, что позволяет учесть все нелинейные закономерности внутри данных.

- › Заказчику передается информация по любому участку в удобной для него форме (набор цифр, который заказчик вставляет в свою модель, не обладая при этом специфическими компетенциями в геоаналитике).
- › Используются специализированные нейросети, которые решают определенный класс задач; они относительно недороги¹⁹⁹ и при этом дают заказчику определенную свободу действий.

Безопасность

Платформа позволяет безопасно объединять данные о клиентах, полностью их обезличивая, но сохраняя ценность для задач моделирования. Она не работает с чувствительными данными. Платформа не делает разовые срезы данных, идет постоянный мониторинг, модели постоянно обновляются. За соблюдение требований информационной безопасности отвечает специальное подразделение банка.

Авторы раздела:



В. В. Малыгин



М. В. Туманова



А. А. Орлова

Б5 Использование данных мобильного оператора Tele2

Название проекта Геоаналитика сотового оператора для госслужбы
Год 2019–2021

Название организации BigData Office Tele2

Цель проекта Оптимизация транспортных потоков, решение градостроительных задач, обеспечение безопасности на массовых мероприятиях, уточнение демографических данных.

Команда Дата-аналитики, дата-инженеры, исследователи данных.

Результат Созданы аналитические инструменты для принятия управленческих решений. Получена новая оценка данных для улучшения качества процессов.

Интересный факт В 2014 году на основе геоданных аналитики Tele2 было установлено, что на одном из московских кладбищ постоянно проживают около 2500 человек. Проверка выявила общежитие нелегальных мигрантов.

О проекте

Технология обработки данных мобильного оператора, разработанная в подразделении больших данных компании Tele2, использовалась для решения задач в сфере туризма, транспорта, обеспечения безопасности на массовых мероприятиях, муниципального управления; определения количества и качества населения, изучения пропускной способности дорог. Сотрудники Tele2 изучали запросы РОИВ и помогали решать им демографические, транспортные и другие проблемы (см. табл. 9).

¹⁹⁹ По сравнению с универсальной сетью.

Таблица 9. Что дает использование мобильных данных

Сфера	Назначение
Туризм	Позволяет составить портрет каждого туриста: пол, возраст, откуда он приехал, на каком транспорте, чем интересовался, когда и как уехал обратно и т. п.
Безопасность на массовых мероприятиях	Мобильные данные позволяют не только установить точное число участников мероприятия, но и составить их портрет. Это помогает планировать мероприятия с учетом потребностей их участников
Городской пассажирский транспорт	Аналитика на основе мобильных данных дает городским властям понимание, насколько оптимальна действующая транспортная сеть, какие маршруты пользуются популярностью, сколько людей ими пользуется
Демография	Геоданные мобильного оператора позволяют получить точную картину распределения населения на площади в конкретный момент времени

1. Туризм

Комитет по развитию туризма **Волгоградской области** попросил сотрудников Tele2 проанализировать состав туристов, посещающих регион. Госслужащие считали, что к ним едут преимущественно родственники волгоградцев, а также рыбаки или туристы. Анализ турпотока с помощью мобильных данных показал, что через регион перемещается много транзитных туристов. Среди них выделяется особый сегмент — байкеры, которые выбирают вместо федеральных трасс дороги с более спокойным трафиком и останавливаются в популярных в их среде местах по дороге на байк-фестивали. Региональные власти оборудовали в этих местах стоянки, сделали кемпинги, поставили туалеты, там стала развиваться коммерческая инфраструктура.

Один из курортов регулярно отчитывался о росте туристического потока. Однако при анализе данных пользователей сотовой сети (см. рис. 39) специалисты Tele2 выяснили, что 95% приезжих раньше уже бывали на



При перемещении телефон абонента переключается между базовыми станциями



Оператор связи видит данные о нагрузке на базовые станции и хранит их



Через агрегацию данных абонентов в конкретной точке формируются потоки информации



Телеком-оператор — единственный, кто видит поведение своего абонента 24x7, а постоянный мониторинг долей рынка позволяет экстраполировать оценки на все население



Анализ актуальной информации, основанной на фактическом поведении туристов, вместе с ретроспективными данными как по поведению, так и по интересам



Анализ местоположения, включая определение зон проживания, работы, отдыха и развлечений



Анализ миграции, используемого транспорта, помощь в оптимизации социальной инфраструктуры

Рисунок 39. Использование данных оператора связи для анализа турпотоков

этом курорте, новых туристов практически нет. Эта проблема ставила под сомнение перспективы курорта, и ее невозможно было выявить традиционными методами, при которых считали количество посещений, но не учитывали уникальность персон.

Для учета туристов, прибывающих в район популярного природного парка **Сортавала**, администрация Республики Карелия традиционно запрашивала сведения у гостиниц, РЖД и Росавтодора, а также у ФСБ (о приехавших иностранцах). На основе этих данных госслужащие делали выводы об объеме турпотока и необходимой инфраструктуре. Когда специалисты Tele2 показали администрации региона свои данные, основанные на геолокации сотовых телефонов, оказалось, что количество туристов почти в 10 раз превышает цифры, полученные в результате обычного подсчета: на обсуждаемой территории находилось более 1 млн гостей; гостиницы были переполнены.

Каким образом были получены эти данные? Требовалось определить домашнюю локацию абонента (место, где он в основном находится в ночное время). Система устанавливает домашнюю локацию на основе динамических изменений, после этого можно выделить туристов. Оператор мобильной связи обрабатывает петабайты информации, но вся она обезличена, результаты формируются в агрегированном виде.

Специалисты Tele2 взаимодействовали с исследовательским агентством, которое разрабатывало стратегию развития туризма в регионе. В проекте был проанализирован турпоток за два предыдущих года.

2. Демография

Для развития населенного пункта важно знать, сколько человек проживают там постоянно, сколько — временно, сколько человек посещают город в определенные часы. Раньше использовались два способа: регистрация по месту жительства (данные МВД) и перепись населения (Росстат), — но перепись дает устаревшие и неточные данные, а регистрация не учитывает суточную мобильность (рис. 40).



Рисунок 40. Расчет суточной мобильности

Тепловые карты, построенные на основе мобильных данных, позволяют увидеть городские проблемы, от скопления мусора до неудобного маршрута транспорта и недостатка социальной инфраструктуры, а также наблюдать за активностью населения, например подсчитать, сколько человек регулярно совершают пробежки.

В деревне под **Ижевском** рассматривался вопрос о необходимости фельдшерского пункта. Команда Tele2 провела анализ годовой динамики населения, и стало понятно, что это скорее дачный поселок: летом туда приезжают отдыхать, а зимой там почти никто не живет. График работы фельдшерского пункта был изменен, и теперь он работает с мая по октябрь, во время дачного сезона.

3. Безопасность на массовых мероприятиях

Специалисты Tele2 сотрудничали с администрацией **Санкт-Петербурга**, помогая составить социально-демографический портрет участников массовых мероприятий. На одном из фестивалей организаторы неожиданно для себя обнаружили большое количество детей. Анализ данных сотовых операторов показал, что дети приехали из Челябинской области, где каникулы начались на две недели раньше. В точной характеристике посетителей мероприятий был заинтересован не только городской комитет по развитию туризма, но и комитет по культуре: если первый отвечал за приезжих, то второму было важно понять, сколько петербуржцев посещает такие фестивали.

Анализ данных помогает планировать мероприятия с учетом потребностей их участников. В итоге было изменено время работы рамок безопасности, оптимизированы маршруты подвоза туристов, скорректирована маркетинговая активность (исходя из портрета посетителя).

4. Городской транспорт

При строительстве объездной дороги вокруг **Сочи** необходимо было рассчитать ее пропускную способность. Проектная организация, агентство «Стройпроект», собиралась применить традиционные методы, чтобы рассчитать усредненный транспортный поток за день, неделю или месяц. Такой способ не позволял точно прогнозировать пиковые нагрузки и не давал возможности избежать пробок в часы пик. Для анализа транспортной нагрузки команда Tele2 использовала матрицы корреспонденций (перемещений из одного места в другое) за получасовые интервалы, при этом учитывался план строительства новых жилых районов на пути трассы. В результате пропускная способность дороги была рассчитана на несколько лет вперед и с учетом повышенной нагрузки в часы пик.

В **Иркутске** городские власти закрывали для ремонта дороги одну из улиц и попросили команду Tele2 рассчитать, насколько труднее будет людям передвигаться по городу. По результатам аналитики время перемещения человека из дома на работу должно было увеличиться в полтора-два

раза. Такие точные подсчеты помогают выбрать правильное решение при планировании работ.

Раньше количество пассажиров вручную фиксировали специальные сотрудники в течение нескольких дней. Сейчас аналитик за несколько часов может построить матрицы пассажиропотока в общественном транспорте с помощью данных мобильных операторов. Специалисты Tele2 делят город на транспортные районы исходя из доступности рабочих мест, социальных объектов по отношению к местам проживания, составляют матрицы корреспонденций между районами, которые показывают миграционную потребность населения — транспортный спрос, затем на матрицу корреспонденций накладывается существующая маршрутная сеть и генерируются передвижения населения, на основе которых можно сделать вывод о загруженности маршрута и востребованности направления и о подвижности населения в целом.

В частности, можно рассчитать для каждого жителя города расстояние до парка (в проектах по развитию городского хозяйства есть показатель — пешая доступность парка для каждого горожанина); по итогам анализа можно спроектировать зеленые зоны.

Технологии

Инструменты работы с данными показаны в табл. 10.

Результаты

Работа с данными позволила, в частности:

- › уточнить объем туристического потока, который оказался в 10 раз больше предполагаемого;
- › обнаружить поток транзитных туристов на автотрассе и оборудовать для них стоянки;

Таблица 10. Инструменты работы с данными

Инструмент	Применение
Персонализация и лидогенерация	Определение туристического потока
Транспортное моделирование	Определение пропускной способности дорог
Распределение населения	Определение количества и качества населения
Тепловые карты территорий и геониты	Определение количества и качества населения
Нейронная сеть	Прогнозирование
Цифровые двойники	Дистанционное зондирование земли
Математическая экстраполяция	Во всех проектах

- › установить регионы, откуда прибыли участники массовых мероприятий, оптимизировать туристические маршруты;
- › повысить точность долгосрочного планирования и избежать транспортных проблем в перспективе;
- › проанализировать причины ДТП, чтобы обеспечить безопасность движения.

На основе проделанной работы был сделан вывод, что для обоснованного прогноза нет необходимости получать данные от всех сотовых операторов: экстраполяция данных одного-двух операторов дает прогноз необходимой точности. Но получить нужные данные недостаточно, нужна готовность их использовать: вносить изменения в проекты строительства, в планы ремонта, в обеспеченность социальной инфраструктурой.

Б6 Система раннего предупреждения онкозаболеваний в Якутии

Авторы раздела:



Л. Н. Афанасьева



А. А. Орлова

Название проекта Создание веб-приложения «Онкопоиск»

Год 2020

Название организации Министерство здравоохранения Республики Саха (Якутия), Якутский республиканский онкологический диспансер

Цели проекта Ранняя диагностика злокачественных новообразований и предопухоловой патологии, выявление шести видов рака, наиболее распространенных в Якутии.

Формирование групп повышенного онкологического риска и отбор нуждающихся в дополнительном обследовании.

Выявление людей с высоким риском возникновения онкозаболеваний.

Команда

- › Авторы программы: главврач и врачи Якутского республиканского онкодиспансера.
- › Разработчики программы: ИТ-специалисты.
- › Для разработки сервиса был назначен CDO.

Результат Программа находится в стадии внедрения, планируется подключение медицинских организаций первичного звена. В дальнейшем веб-приложение сможет использовать любой житель республики.

Интересный факт Кроме онлайн-диагностики проект «Онкопоиск» предполагает также открытие кабинетов онкоскрининга в поликлиниках и выезд в отдаленные районы мультидисциплинарных медицинских бригад «Онкодесант».

О проекте

«Онкопоиск» нацелен на выявление шести наиболее распространенных в регионе видов рака. Представляет собой автономный сайт-платформу для автоматизированного расчета рисков развития злокачественных новообразований (ЗНО) и предраковых состояний.

Таблица 11. Архитектура проекта «Онкопоиск»

Часть системы	Пользователь	Функции	Цель
Пользователь-ская	Пациент	Регистрируется в системе, отвечает онлайн на вопросы анкеты	Возможность записи на прием и доступ к личному кабинету, где хранятся результаты обследований
Врачебная	Врач	Интерпретирует результаты обследования, формирует группу пациентов для дополнительного обследования, беседует с пациентом	Организация дообследования при выявлении подозрения на злокачественное новообразование
Административная	Регистратор	Фиксирует движение участников системы онлайн, подтверждает запись на обследование	Маршрутизация, заполнение медицинской документации
Аналитическая	Аналитик	Создает отчеты на основе данных	Анализ работы системы

Возможности программы «Онкопоиск»²⁰⁰ (см. табл. 11):

- › Анкетирование и диагностика. Алгоритм анализирует результаты заполнения опросника и выбирает пациентов, которым нужно дополнительное обследование в онкодиспансере. Все пользователи системы фиксируются в республиканском регистре скрининга рака.
- › Аналитика. С помощью программы можно вести онлайн-мониторинг, а также составить карту заболеваемости по республике.

Предпосылки

Необходимость в скрининговом сервисе назрела давно; ранее подобные анкеты были размещены на сайте республиканского онкодиспансера, но граждане редко заполняли их. При подготовке проекта по данным формы № 7 федерального статистического наблюдения были изучены региональные особенности заболеваемости и смертности в Якутии от злокачественных новообразований за последние 10 лет. Оказалось, что шесть заболеваний в сумме давали более 50% всей смертности от ЗНО. Команда проекта выбрала именно эти заболевания. По ним имелись скрининговые технологии с доказанной эффективностью. В качестве основы были взяты анкеты, разработанные ВОЗ, и адаптированы с учетом региональной специфики (заболеваемости и смертности), добавлены вопросы по выявлению малых симптомов рака по шести нозологиям.

Технологии

Команда разработки изучила технологии, позволяющие быстро создавать безопасные и поддерживаемые веб-сайты, и создала на основе сочетания фреймворков (Django, FastAPI, Vue, React.js) программно-аналитический

²⁰⁰ Скрининг и ранняя диагностика рака в Якутии. URL: <https://онкопоискаха.рф>

комплекс. Обработка и визуализация данных происходят в системе Yandex Datalens. Основной технической сложностью стала интеграция с существующими медицинскими системами.



Пилотный проект (скрининг рака легких) реализован в 2020 году в Якутске, в 2021 году добавлен скрининг остальных видов рака. Когда количество анкет пациентов, имеющих повышенный риск развития онкозаболеваний, достигнет 25 000, будет оценена эффективность тестирования.

Чтобы вовлечь в анкетирование больше участников, скрининг-платформу активно продвигали на популярных в регионе информационных каналах (на официальных ресурсах, в социальных сетях). Использовалось сотрудничество с руководством разного уровня — от республиканского до муниципального, анонс сервиса перевели на якутский язык, учли возможность распространения информации от пользователя к пользователю по WhatsApp.

Безопасность

Были подписаны договоры о сотрудничестве с Якутским региональным онкодиспансером и Минздравом, использовались https-протоколы шифрования. Компания, которая обеспечивает кибербезопасность Якутского республиканского онкодиспансера, протестировала сервер и устройства, на которых они обрабатывались. Сервер аккредитован в соответствии с федеральным законом «О персональных данных» .

Б7 Повышение эффективности управления филиалами МФЦ в Новосибирской области

Название проекта Разработка и внедрение ИС МФЦ

Год 2018

Название организации Министерство цифрового развития и связи Новосибирской области

Цель проекта Повышение эффективности и качества предоставления услуг многофункциональных центров Новосибирской области.

Команда Представители министерства цифрового развития и связи Новосибирской области, руководители МФЦ и разработчики основного платформенного решения для МФЦ, на базе которого реализован модуль бизнес-аналитики.

Результат Среднее время ожидания в очереди сократилось на 23%. Время сбора квартальной отчетности сократилось по всей сети МФЦ с пяти дней до одной минуты.

Авторы раздела:



О. С. Жданова



А. А. Орлова

Интересный факт В результате внедрения системы повысилась скорость работы операторов; суммарное время приема граждан по всей сети МФЦ сократилось на 870 часов в месяц.

О проекте

Система бизнес-аналитики МФЦ позволяет наблюдать работу МФЦ в реальном времени и автоматически определяет отклонения от средних значений, например:

- › срок ожидания свыше 15 минут;
- › сроки часто нарушаются;
- › недостаточное количество операторов;
- › неравномерность загрузки окон в филиале;
- › неравномерность загрузки филиалов;
- › увеличение среднего времени обслуживания в филиале.

Информация об отклонениях передается всем заинтересованным лицам: директорам филиалов, руководству сети, контролирующим органам.

В интерфейсе системы бизнес-аналитики доступен просмотр и анализ всех важных параметров работы МФЦ:

- › количества заявлений;
- › количества консультаций;
- › количества выданных документов;
- › среднего времени ожидания;
- › среднего времени обслуживания оператора по различным услугам;
- › количества ошибок операторов;
- › оценок системы «Ваш контроль»;
- › статистики работы контакт-центра и др.

Особенности проекта

Для каждого оператора по каждой услуге и виду приема (первичный прием документов, выдача готовых результатов) рассчитывается среднее время обслуживания. Это время сравнивается со средним временем обслуживания по этой услуге и виду приема по всей сети. Если показатель оператора более чем на 50% превышает средний показатель по сети, оператору выставляется «отклонение» по этой услуге и виду приема.

Затем руководитель оператора изучает статистику системы аналитики и помогает оператору повысить показатели. Если непосредственный руководитель в течение трех месяцев не смог помочь оператору устранить отклонение, проблема переадресуется вышестоящему руководству.

Для повышения эффективности в МФЦ используют систему мотивации. За каждое принятое заявление оператору начисляют баллы (с учетом скорости

Таблица 12. Проблемы в работе МФЦ Новосибирской области и их решения, основанные на системе бизнес-аналитики

Проблема	Решение
Построение аналитической отчетности может увеличить нагрузку на основные системы оперативного учета (АИС МФЦ, электронную очередь, контакт-центр)	Данные различных систем размещены в едином хранилище
Талоны электронной очереди использовались не всегда, это приводило к искажению данных о времени ожидания	Административные меры: корректировка регламентов, работа с персоналом; программные решения: предупреждение в случае обслуживания без талона, отчетность по фактам нарушений
Некоторым руководителям филиалов было сложно принять новую систему мотивации сотрудников, в частности распределения премий	Поэтапный ввод системы мотивации. На первом этапе руководители имели право назначать премии по своему усмотрению, но с обязательным указанием количества баллов, набранных оператором. В случае вопросов к оценкам обсуждали корректность формул расчета баллов, руководители вносили предложения. В результате доработок была выведена формула расчета, которая устраивала как руководство сети МФЦ, так и большинство руководителей филиалов

обслуживания, количества ошибок, универсальности оператора и оценок, выставленных заявителями в системе «Ваш контроль»). Используется как нематериальная (соревнование сотрудников, доска почета, грамоты), так и материальная мотивация (премиальный фонд распределяется между операторами в соответствии с набранным количеством баллов).

Способы решения ряда проблем, выявленных в ходе проекта, представлены в таблице 12.

Нормативные документы

Разработка и внедрение информационной системы шли в соответствии с требованиями законодательства о размещении госзаказов. Техническое задание на разработку системы формировалось совместно специалистами министерства цифрового развития и связи Новосибирской области и сотрудниками МФЦ Новосибирской области.

Технологии

Источники данных: автоматизированная информационная система (АИС) МФЦ, электронная очередь, контакт-центр, 1С. Данные обрабатываются и помещаются в хранилище. На основании анализа данных система формирует набор показателей, интерактивных панелей, отчетов.

Безопасность

Анализируемые данные не являются персональными или конфиденциальными, не составляют государственной тайны. Используются стандартные методы обеспечения ИБ.



В. ПРАВОВЫЕ ОСНОВЫ УПРАВЛЕНИЯ ГОСДАНЫМИ

 *Время чтения: 19 мин.*

Принципы и механизм управления государственными данными обозначены в основных стратегических документах Российской Федерации. В последнее время управление данными становится предметом еще и нормативно-правового регулирования. Содержательные аспекты работы с данными подчинены прежде всего требованиям информационного законодательства: в нем установлены правовые режимы информации, на основе которых выделяются правовые режимы данных. Существование различных правовых режимов данных обеспечивает их надлежащую обработку и полноценное участие в информационном обмене.

В1 Система управления данными — основа цифровизации госуправления

Создание полноценной системы управления качеством данных госсектора названо в **документах стратегического планирования страны** необходимым условием достижения целей социально-экономического

развития. В Стратегии развития информационного общества в РФ на 2017–2030 годы предусмотрено совершенствование регулирования в сфере обеспечения безопасной обработки информации и применения новых технологий, упорядочение алгоритмов обработки данных и доступа к таким данным²⁰¹. Важность доступа к наборам данных, особенно обезличенным, обозначена в Стратегии развития ИИ²⁰². Для достижения национальных целей правительству поручено обеспечить создание инфраструктуры высокоскоростной передачи и обработки больших объемов данных, доступной для всех организаций и домохозяйств²⁰³.

Меры по повышению эффективности управления данными рассмотрены и в отраслевых проектах. Согласно федеральному проекту «Цифровое государственное управление»²⁰⁴ госорганы обязаны внедрять разного рода цифровые платформы в целях организации межведомственного взаимодействия, облегчения предоставления услуг, обеспечения прав граждан. В связи с этим на базе системы госуслуг в 2020 году была запущена платформа «Цифровой профиль гражданина», которая становится единым цифровым «окном» для взаимодействия гражданина с государством и бизнесом²⁰⁵.

До 2024 года планируется интеграция пространственных данных из регионов в Единую цифровую платформу²⁰⁶, запланировано²⁰⁷ создание к 2024 году генеральной схемы развития центров обработки данных. Также утверждена Концепция создания и функционирования национальной системы управления данными²⁰⁸. НСУД (см. раздел 3.1.2) закрепляет единые для всех ГИС стандарты качества данных и требования к управлению ими. Чтобы сделать госданные более доступными, проводится эксперимент по использованию компонентов НСУД для обмена сведениями в электронной форме между физическими лицами, банками, госорганами²⁰⁹.

Авторы:



М. А. Ильященко



Д. А. Кзылходжаева



В. А. Сохоров

²⁰¹ Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001201705100002>

²⁰² Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001201910110003>

²⁰³ Паспорт национального проекта «Национальная программа „Цифровая экономика Российской Федерации“» // Министерство цифрового развития, связи и массовых коммуникаций РФ. URL: <https://digital.gov.ru/ru/activity/directions/858>

²⁰⁴ Паспорт федерального проекта «Цифровое государственное управление». URL: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Pasport-federalnogo-proekta-Cifrovое-gosudarstvenное-управление.pdf>

²⁰⁵ Тер: цифровой профиль // D-Russia.ru. URL: <https://d-russia.ru/tag/tsifrovoy-profil/>; 3,5 млн согласий на предоставление различных сведений дано за год работы цифрового профиля // Министерство цифрового развития, связи и массовых коммуникаций РФ. URL: <https://digital.gov.ru/ru/events/40998/>

²⁰⁶ Постановление Правительства РФ от 01.12.2021 № 2148 «Об утверждении государственной программы Российской Федерации «Национальная система пространственных данных»» // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001202112070014>

²⁰⁷ Паспорт федерального проекта «Информационная инфраструктура». URL: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Pasport-federalnogo-proekta-Inформационная-инфраструктура.pdf>

²⁰⁸ Распоряжение Правительства РФ от 03.06.2019 № 1189-р «Об утверждении Концепции создания и функционирования национальной системы управления данными и плана мероприятий („дорожной карты“) по созданию национальной системы управления данными на 2019–2021 годы» // ГАРАНТ.РУ. URL: <https://www.garant.ru/products/ipo/prime/doc/72162090/>

²⁰⁹ Постановление Правительства РФ от 17.08.2021 № 1361 «О внесении изменений в постановление Правительства Российской Федерации от 3 июня 2019 г. № 710» // Правительство России. URL: <http://government.ru/docs/all/136124/>

В2 Классификация государственных данных

В российском законодательстве понятие «данные» как отдельная правовая категория не выделяется, а включено в качестве одной из характеристик более общего понятия — «информация»²¹⁰ (о связи этих терминов см. раздел 2.1). Поэтому и правовой статус данных регулируется правовым режимом информации, который установлен исходя из целей сбора и назначения данных. При этом технические параметры «данных» сформулированы в государственных стандартах²¹¹.

Основы регулирования и разграничения информации были определены в **Федеральном законе от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации»**. В соответствии с ним органы власти в рамках своей компетенции реализуют права и обязанности обладателя (или оператора) информации от имени Российской Федерации, того или иного субъекта РФ или муниципального образования. Госорганы — не собственники, а администраторы собираемых данных; они используют данные (в том числе совместно с другими уполномоченными органами) для достижения целей социально-экономического развития страны, региона, муниципалитета.

По уровню доступа поступающие и хранящиеся государственные и муниципальные данные подразделяются на **общедоступную информацию** и **информацию ограниченного доступа**. Исходя из принципа гласности государственного и муниципального управления по общему правилу информация, создаваемая органом власти, является общедоступной. Общедоступная информация может использоваться любыми лицами по их усмотрению. На другом конце спектра находятся конфиденциальные сведения, отнесенные к информации ограниченного доступа на основании федерального закона (см. подробнее в разделе В4). Это, в первую очередь, персональные данные физических лиц, а также информация, составляющая государственную²¹², служебную²¹³ и другие виды тайн. Следует отметить сложность разграничения информации ограниченного доступа: нередко каждая категория данных имеет несколько подкатегорий со своими особенностями, а сами границы могут быть нечетко очерчены законодателем или изначально являются динамичными.

Учитывая значимость данных частных субъектов для создания «больших данных» следует особо выделить категорию персональных данных (о них

²¹⁰ П. 1 ст. 2 Федерального закона от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации». URL: <http://pravo.gov.ru/proxy/ips/?docbody&nd=102108264>

²¹¹ ГОСТ Р ИСО/МЭК 20546–2021. Национальный стандарт Российской Федерации. Информационные технологии. БОЛЬШИЕ ДАННЫЕ. Обзор и словарь // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/1200180276>

²¹² Закон РФ от 21.07.1993 № 5485–1 «О государственной тайне». URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102025035>

²¹³ Постановление Правительства РФ от 03.11.1994 № 1233 «Об утверждении Положения о порядке обращения со служебной информацией ограниченного распространения в федеральных органах исполнительной власти». URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102099047>

и о проблемах работы с ними см. раздел 3.2.1). В соответствии с **Федеральным законом от 27.07.2006 № 152-ФЗ «О персональных данных»** (далее — ФЗ-152) персональные данные — это любая информация, которая прямо или косвенно позволяет определить физическое лицо (или сделать его определяемым по тем или иным признакам). То есть персональные данные — это не только паспортные данные, но и другая информация о человеке (например, уровень доходов, место проживания). Кроме того, особо чувствительные сведения (например, о состоянии здоровья) относятся к специальной категории персональных данных.

Согласно ФЗ-152, обработка персональных данных должна осуществляться на законной и справедливой основе, то есть для достижения конкретной, заранее определенной цели в интересах субъекта персональных данных, после чего эти данные подлежат уничтожению. По общему правилу обработке персональных данных должно предшествовать получение согласия субъекта персональных данных на такую обработку. Однако для госструктур установлено исключение из этого правила для предоставления государственных и муниципальных услуг, судопроизводства и в некоторых иных случаях. Также не требуется согласие субъекта, если данные предоставляются в исследовательских целях в обезличенном виде согласно требованиям, утвержденным Приказом Роскомнадзора от 05.09.2013 № 996 «Об утверждении требований и методов по обезличиванию персональных данных».

В3 Требования к защите данных и информационному обмену

Установление правового статуса данных необходимо для определения мер их защиты, а также вовлечения этих данных в информационный обмен. С точки зрения внутренних процессов к государственным данным предъявляются повышенные требования даже при обработке общедоступной информации. Например, государственные информационные ресурсы чаще всего используются в формате ГИС, для которых установлены специальные требования к порядку их эксплуатации²¹⁴. В свою очередь, ГИС должны соответствовать определенным требованиям информационной защиты даже при хранении данных, не относящихся к государственной тайне²¹⁵.

В качестве операторов персональных данных госструктуры должны обеспечивать конфиденциальность персональных данных (не раскрывать их другим лицам, кроме прямо предусмотренных законом случаев), а также принимать специальные организационно-технические меры,

²¹⁴ Постановление Правительства РФ от 06.07.2015 № 676 «О требованиях к порядку создания, развития, ввода в эксплуатацию, эксплуатации и вывода из эксплуатации государственных информационных систем и дальнейшего хранения содержащейся в их базах данных информации». URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102375086>

²¹⁵ Приказ ФСТЭК России от 11.02.2013 № 17 // ФСТЭК России. URL: <https://fstec.ru/normotvorcheskaya/akty/53-prikazy/702>; Приказ ФСБ РФ и ФСТЭК России от 31.08.2010 № 416/489 // ГАРАНТ.РУ. URL: <https://base.garant.ru/199541/>

направленные на обеспечение и защиту обрабатываемых персональных данных. Эти меры перечислены в Постановлении Правительства РФ от 21.03.2012 № 211²¹⁶, а также в подзаконных актах ФСТЭК России²¹⁷.

Если же говорить о взаимодействии за пределами контура госоргана, то законодательство накладывает на структуры государственного и муниципального звена значительные требования по обязательному раскрытию информации, в том числе а) в виде открытых данных и б) через систему межведомственного электронного взаимодействия.

а) **Федеральный закон от 09.02.2009 № 8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»** (далее — ФЗ-8) устанавливает перечень и структуру тех данных, которые обязаны обнародовать государственные органы и органы местного самоуправления. Общедоступная информация о деятельности госорганов и органов местного самоуправления предоставляется неограниченному кругу лиц в форме открытых данных. Открытые данные — это информация, размещаемая ее обладателями в сети «Интернет» в формате, допускающем автоматизированную обработку (см. подробнее в разделе 3.1.3).

В развитие ФЗ-8 было принято Постановление Правительства РФ «Об обеспечении доступа к информации о деятельности Правительства Российской Федерации и федеральных органов исполнительной власти»²¹⁸. Требования к порядку размещения открытых данных установлены также Распоряжением Правительства РФ от 10.07.2013 № 1187-р²¹⁹ и Приказом Минкомсвязи России от 27.06.2013 № 149²²⁰.

б) **В Федеральном законе № 210-ФЗ от 27.07.2010 «Об организации предоставления государственных и муниципальных услуг»** закреплён механизм межведомственного взаимодействия при предоставлении этих услуг, соответствующий административным регламентам.

Следует отметить, что в соответствии с Постановлением Правительства РФ от 20.07.2021 № 1228²²¹ к 31 декабря 2022 года все административные

²¹⁶ Постановление Правительства РФ от 21.03.2012 № 211 «Об утверждении перечня мер, направленных на обеспечение выполнения обязанностей, предусмотренных Федеральным законом „О персональных данных“ и принятыми в соответствии с ним нормативными правовыми актами, операторами, являющимися государственными или муниципальными органами» // Правительство России. URL: <http://government.ru/docs/all/81438/>

²¹⁷ Приказ ФСТЭК России от 18.02.2013 № 21 «Об утверждении Составы и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных» // КонсультантПлюс. URL: http://www.consultant.ru/document/cons_doc_LAW_146520/

²¹⁸ Постановление Правительства РФ от 24.11.2009 № 953 // Правительство России. URL: <http://government.ru/docs/all/70318/>

²¹⁹ Распоряжение Правительства РФ от 10.07.2013 № 1187-р «О перечнях информации о деятельности государственных органов, органов местного самоуправления, размещаемой в сети „Интернет“ в форме открытых данных» // Правительство России. URL: <http://government.ru/docs/3243/>

²²⁰ Приказ Минкомсвязи России от 27.06.2013 № 149 «Об утверждении Требований к технологическим, программным и лингвистическим средствам, необходимым для размещения информации государственными органами и органами местного самоуправления в сети „Интернет“ в форме открытых данных, а также для обеспечения ее использования» // Гарант. URL: <https://base.garant.ru/70438250/>

²²¹ Постановление Правительства Российской Федерации от 20.07.2021 № 1228 // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001202107260032>

регламенты государственных услуг федерального уровня должны быть переведены в цифровой (машиночитаемый) формат. Цифровые регламенты будут формироваться с помощью конструктора стандартных текстовых блоков на базе ФГИС «Федеральный реестр государственных и муниципальных услуг (функций)». Ожидается, что это нововведение поможет более подробно описать «клиентский путь» пользователя государственных услуг, а значит — индивидуализировать процесс оказания услуг, сделать его проактивным.

В4 Проверка правового режима данных: чек-лист

1. Для эффективной работы с данными целесообразно определиться с перечнем действующих **нормативных правовых актов**, которые ограничивают возможность распространения данных (федеральные, региональные законы и подзаконные акты, ведомственные правовые акты).

2. Имеющиеся в активе данные необходимо **инвентаризировать** не только с учетом источников их поступления, но и подразделяя на классификационные категории — **структурировать данные по степени ограничений их распространения** (здесь пригодится ранее сформированный перечень НПА, не придется делать двойную работу).

Информация подразделяется на конфиденциальную (ограниченная по распространению или запрещенная в соответствии с законодательством РФ) и общедоступную. К информации, ограниченной к распространению или запрещенной, относится та, распространение которой регулируется законодательно:

- › сведения, составляющие государственную тайну (их перечень определен Указом Президента РФ от 30.11.1995 № 1203);
- › конфиденциальная информация (служебная, коммерческая, врачебная, банковская, нотариальная, налоговая, аудиторская, адвокатская тайна и т. п.);
- › информация об охраняемых результатах интеллектуальной деятельности;
- › информация, признанная имеющей экстремистский характер;
- › пропаганда наркотиков;
- › пропаганда суицида;
- › продукция порнографического характера, в том числе детская порнография;
- › продукция эротического характера;
- › реклама;
- › предвыборная агитация;

- › нецензурная брань;
- › сведения и материалы, порочащие репутацию граждан и юридических лиц;
- › изображения граждан (в том числе несовершеннолетних).

Очевидно, что ограниченная к распространению информация (из большинства перечисленных выше пунктов) не образуется в деятельности государственных органов, но вполне может содержаться в общедоступных источниках в сети «Интернет» (например, в социальных сетях). Поэтому особенно тщательной проверке и классификации подлежат данные, полученные именно из общедоступных источников. Данные, образующиеся в деятельности государственных органов, также следует разграничивать на общедоступные и ограниченные к распространению (например, информация для служебного пользования).

3. Дополнительно стоит учитывать также **свойства информации**: при достижении определенного объема данных или объединении сетов неконфиденциальных данных в определенных комбинациях информация может становиться конфиденциальной. Например, данные за квартал не являются конфиденциальными, за год — могут стать конфиденциальными; данные по области — неконфиденциальные, по федеральному округу или всей стране — конфиденциальные.

4. В ведомстве рекомендуется определить **структурные подразделения, должностных лиц, ответственных за работу с данными** (например, они непосредственно участвуют в создании и обработке данных, осуществляют мониторинг изменений НПА в части ограничений распространения данных в своей отрасли и т. п.). Работа таких структурных подразделений (должностных лиц) должна строиться в тесном взаимодействии с подразделениями, отвечающими за **информационную безопасность**.

5. Результаты инвентаризации и отнесения данных к определенным категориям позволят произвести **маркировку данных**, имеющих в активе (на уровне метаданных приводится описание дополнительных свойств данных, касающихся возможности и условий их распространения или предоставления). Маркированные данные в дальнейшем позволят оперативно определять, имеет ли информация статус общедоступной или она ограничена к распространению, каковы условия ее распространения или передачи. Это повысит скорость и эффективность принятия решений, связанных с управлением данными.

6. Далее следует произвести **проверку элементов ИТ-инфраструктуры** на предмет соответствия уровня их защищенности и хранящихся и обрабатываемых в них маркированных данных. Результатом такой проверки должно стать распределение и хранение данных в соответствии с их категорией в информационных системах с надлежащим уровнем защищенности.

ИСТОЧНИКИ ИЛЛЮСТРАЦИЙ

Средневековые книжные иллюстрации были заимствованы из следующих открытых источников:

- › Fondation Martin Bodmer (Cologne, Genève), Cod. Bodmer 49. Christine de Pisan, Epitre d'Othea
- › Bibliothèque nationale de France. Christine de Pisan, Cité des dames
- › Bibliothèque nationale de France. Département des Manuscrits. Le secret de l'histoire naturelle contenant les merveilles et choses mémorables du monde
- › The Pierpont Morgan Library. The Morgan bible
- › British Library. Talbot Shrewsbury book
- › British Library. Les grandes chroniques de France

Источники рисунков **1, 3, 9, 18** указаны в сносках к тексту. Схемы и рисунки **2, 5, 10, 14, 17, 25, 26, 28–30, 33, 34, 36–40**, иллюстрирующие содержание доклада, были предоставлены авторами соответствующих разделов. Это как новые авторские материалы, так и адаптированные материалы организаций, в которых они работают. Схемы и рисунки **6, 7, 19, 20, 23** заимствованы из русского издания DAMA-DMBOK, в подготовке которого участвовали авторы доклада (DAMA-DMBOK: свод знаний по управлению данными. 2-е изд. М.: Олимп–Бизнес, 2020).

Остальные рисунки были заимствованы из открытых источников, указанных ниже. В некоторых случаях они были переведены или частично изменены авторами доклада.

- Рис. 4.** Взаимосвязь между ресурсами, способностями и конкурентным преимуществом. — Grant R. M. The resource-based theory of competitive advantage: implications for strategy formulation // California Management Review. 33 (3). Pp. 114–135, Berkeley, Calif.: University of California, 1991.
- Рис. 8.** Пирамида знаний. — Hey J. The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link // OceanTeacher: a training system for ocean data and information management. Intergovernmental Oceanographic Commission (UNESCO), 2004.
- Рис. 11.** Цикл преобразований «данные — информация — знания». — 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.
- Рис. 12.** Взаимосвязи основных категорий данных в деятельности организации. — van Gils B. Data Management: a gentle introduction. Van Haren Publishing, 's-Hertogenbosch, 2020.
- Рис. 13.** Роли отдельных категорий данных в информационном обеспечении процессов организации. — McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. Morgan Kaufmann, 2008.
- Рис. 15.** 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.
- Рис. 16.** Основные компоненты ИСУД. — Единая информационная платформа Национальной системы управления данными. URL: <https://nsud.info.gov.ru>
- Рис. 21.** 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.
- Рис. 22.** Руководство данными и дата-стратегия. — 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.
- Рис. 24.** Пример диаграммы оценки зрелости управления данными. — Medyckyj-Scott D., Stock K., Gibb R., Gahegan M., Dzierzon H., Schmidt J. & Collins A. Our Land and Water National Science Challenge — A Data Ecosystem for Land and Water Data to Achieve the Challenge Mission, 2016. URL: https://www.researchgate.net/figure/The-six-key-themes-of-CMMIs-Data-Management-Maturity-DMM-model-34_fig1_309792608
- Рис. 27.** Экосистема больших данных. — Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2018.
- Рис. 31.** Этапы аналитического проекта. — Runkler T. A. Data Analytics. Models and Algorithms for Intelligent Data Analysis. Springer Vieweg, 2016.
- Рис. 32.** Виды диаграмм и типы сравнения, которые в них используются. — Желязны Д. Говори на языке диаграмм. М.: МИФ, 2011.
- Рис. 35.** «Завод» по получению из данных информации. — Эккерсон У. У. Панели индикаторов как инструмент управления. М.: Альпина бизнес букс, 2007.

ГЛОССАРИЙ

В глоссарии собраны основные термины, использованные в навигаторе, которые могут быть неизвестны читателю, не имевшему дело с данными, или имеют значение, не совпадающее с общепринятым. Слова, выделенные курсивом, сами являются терминами и описаны в глоссарии. Также указаны те разделы доклада, в которых обсуждаемый термин разъясняется, комментируется или иллюстрируется*.

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
1	Агрегированные данные	Агрегаты	Данные, обобщенные (сведенные) по определенным параметрам (интервалу времени, группе клиентов, товаров...)	6.2
2	Аналитические данные		Данные, которые образуются из основных, справочных и транзакционных данных и используются в аналитической деятельности организации	2.2
3	Архитектор данных		Специалист по созданию архитектуры данных. См. Архитектура данных	5.1
4	Архитектура данных		Схема управления информационными активами в соответствии со стратегией организации; содержит стратегические требования к данным и способы удовлетворения этих требований. См. Архитектор данных	1.3.2, 5.1
5	База данных		Совместно управляемый и используемый набор логически связанных данных (и описание этих данных), предназначенный для удовлетворения информационных потребностей организации	6.1
6	Бизнес-аналитика	Business Intelligence, BI	Деятельность, которая делает возможным проведение в организации изменений, полезных для заинтересованных сторон, путем выявления их потребностей и обоснования решений, описывающих пути реализации изменений	9.5.1
7	Бизнес-глоссарий		Набор эталонных терминов и понятий, описывающих деятельность организации. Используется для создания терминологии, концептуальной и логической моделей, реестра правил качества данных, политик, регламентов, процессов и взаимосвязей между ними	7.2
8	Большие данные	Big Data, биг дата	Данные, поступающие с высокой скоростью, для которых характерно большое разнообразие источников и форматов	2.3
9	Валидация данных		Процесс проверки данных на соответствие требованиям, условиям и ограничениям с целью доказать, что все требования к данным удовлетворены	7.1
10	Валидность данных	Валидированность	Соответствие данных заданному уровню правдоподобности и достоверности, то есть корректность описания объектов с помощью этих данных (например, неотрицательный вес, правильный формат даты). Возникает в результате валидации	7.3
11	Визуализация данных		Представление данных с помощью наглядных графических изображений (обычно — разнообразных диаграмм, схем, карт)	9.4
12	Витрина данных		Эталонный слепок данных, готовый для использования внешними и внутренними пользователями. В частности, результат деятельности ведомства, предназначенный для межведомственного обмена	3.1.3

* Глоссарий подготовлен А. В. Ожаровским.

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
13	Гармонизация данных		Процесс и результат использования единых унифицированных справочников и классификаторов, данных из первоисточников, данных с единообразным описанием и единообразным форматом представления	7.1
14	Государственные данные		Данные, содержащиеся в информационных ресурсах органов и организаций государственного сектора, а также в информационных ресурсах, созданных в целях реализации полномочий органов и организаций госсектора	3.1.1
15	Данные		(В узком смысле слова и в контексте «сырые данные») Дискретные, объективные факты или наблюдения, неорганизованные и необработанные, как правило, не передающие никакого конкретного смысла и не имеющие ценности, потому что лишены контекста и интерпретации	2.1
16	Дата-инжиниринг		Область знаний, связанная с разработкой, тестированием и поддержкой инфраструктуры работы с данными: баз данных, хранилищ и систем массовой обработки и подготовки данных. См. Дата-инженер	6.1
17	Дата-инженер	Инженер данных	Специалист по дата-инжинирингу. См. Дата-инжиниринг	5.1, 6.1
18	Дата-стратегия	Data Strategy, data-стратегия, стратегия работы с данными, стратегия данных, стратегия в области данных	Высокоуровневый план использования информации для получения конкурентных преимуществ и реализации целей организации. См. также Стратегия управления данными	4.3
19	Датасет	Набор данных	Структурированный массив данных, пригодных для обучения на них ИИ-решений	Введение
20	Дашборд		Инструмент для визуализации и анализа информации о бизнес-процессах и их эффективности. Как правило, содержит агрегированные и аналитические данные, а также средства визуализации данных	9.4
21	Дескриптивная аналитика	Описательная аналитика	Анализ и визуализация данных для понимания того, что уже произошло или происходит в режиме реального времени	9.2
22	Диагностическая аналитика		Анализ исторических данных с целью обнаружения причинно-следственных связей, закономерностей и корреляций	9.2
23	Доказательная политика	Evidence-based Policy	Государственная политика, основанная на строго установленных объективных доказательствах. См. также Подход, основанный на данных	1.1
24	Жизненный цикл данных	Data Lifecycle, ЖЦ	Цикл работы с данными «от рождения до смерти», который включает планирование, проектирование и обеспечение доступности данных; создание или получение данных; перемещение, преобразование, хранение; также обслуживание данных и предоставление совместного доступа к ним; использование и улучшение (повышение ценности) данных; ликвидацию данных	4.1
25	Загрузка данных		Процесс внесения данных в какую-либо ИС или базу данных	6.2
26	Знания		Совокупность данных и информации (к которым добавляются экспертные мнения, навыки, опыт, другие знания), которая была организована и обработана с целью передачи понимания, накопленных результатов обучения и компетенции так, чтобы получился ценный актив, который можно применить в текущей деятельности для принятия решений	2.1

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
27	Извлечение данных		Этап работы с данными, на котором данные извлекаются из одного или нескольких источников и подготавливаются к преобразованию. Также процесс получения данных из одного или нескольких источников	6.2
28	Интеграция данных		Процесс объединения данных из различных источников для получения их согласованного представления, в широком смысле — процесс организации регулярного обмена данными между различными ИС предприятия	6.1
29	Интеллектуальный анализ данных	Data Mining	Направление аналитики, которое применяется к очень большим объемам данных. В этом случае вариантов их интерпретации слишком много и перечислить все гипотезы невозможно. Задача в таком случае — извлекать из данных догадки, находить неожиданные взаимосвязи и скрытые закономерности	9.3
30	Информационная система	ИС	Совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств, которые дают возможность пользователям получать те или иные информационные сервисы для выполнения своих задач и функций	3.2.1
31	Информационный ресурс		Данные в любом виде, которые можно многократно использовать для решения задач заинтересованных сторон	1.2
32	Информация		Форматированные данные, обработанные с определенной целью, которым придан смысл посредством добавления контекста	2.1
33	Исследователь данных	Дата-сайентист, Data Scientist, специалист по Data Science	Специалист, который применяет методы науки о данных для анализа больших данных, строит и тестирует математические модели поведения данных. См. Наука о данных	6.3
34	Источники данных		Многообразие того, откуда может поступать информация и данные: различные приложения, подключаемые по API, датчики, внешние устройства и другие источники данных, которые подключаются напрямую или с помощью сетевых протоколов. Также данные могут поступать из систем оперативной обработки транзакций	6.2
35	Карта данных		Графическое представление метаданных и зависимостей между ними, визуализирующее происхождение и характер трансформации при обработке данных	3.2.2
36	Каталог данных		Информационная система для выявления, обработки, хранения метаданных и построения взаимосвязей между ними	7.2
37	Качество данных	Data Quality	Степень, в которой данные удовлетворяют заявленные и подразумеваемые потребности различных заинтересованных сторон (в частности, потребителей данных) и которая позволяет оценить пригодность данных для практического использования	5.1, 7
38	Комитет по данным	Совет по данным	Орган, в который входят руководители основных направлений организации. Его главная задача — задавать принципы и правила, выступать связующим звеном между производителями (поставщиками) и потребителями данных	5.1

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
39	Конвейер данных	Конвейер обработки данных, конвейерная обработка данных, цепочечная схема обработки данных, Data Pipeline	Набор блоков обработки данных, соединенных последовательно, где выход одного блока является входом следующего. Блоки конвейера часто выполняются параллельно или с разделением по времени. Состоит из трех основных элементов: источника, блока(-ов) обработки, пункта назначения	6.1
40	Критичность данных	Чувствительность, сенситивность	Важность данных, ценность, которая определяется субъектом, потребляющим или обрабатывающим данные	8.1
41	Маска ввода		Строка символов, указывающая формат допустимых значений данных при ручном вводе, которую используют для обеспечения единого формата данных, например для полей с телефонными номерами из девяти цифр	7.3
42	Мастер-данные	Основные данные	Данные об объектах и бизнес-сущностях, представляющих ценность для организации (о клиентах, продуктах, работниках, технологиях и материалах и т. п.)	2.2
43	Машинные данные		Информация, автоматически генерируемая компьютером, процессом, приложением или устройством без вмешательства человека	2.2
44	Метаданные		Данные, описывающие структуру и характеристики других данных	2.2
45	Модель данных		Документ (и/или схема), включающий наименования элементов данных, подробные определения данных и метаданных, концептуальные и логические сущности и связи между ними, а также бизнес-правила	4.1
46	Наука о данных	Data Science	Область знания, тесно связанная с интеллектуальным анализом данных, предполагающая, как правило, использование машинного обучения, больших массивов данных и мощных вычислительных средств. Для такого анализа разрабатываются алгоритмы, имитирующие некоторые когнитивные способности человека. См. Исследователь данных	6.3
47	Неструктурированные данные		Данные, произвольные по форме, не имеющие строго определенной структуры и не организованные по определенным правилам	2.2
48	Нормализация данных		Процесс приведения структуры данных к требуемому виду. См. также гармонизация	6.2
49	Обогащение данных		Процесс насыщения данных новой информацией (например, используя дополнительные источники данных), которая позволяет сделать их более ценными с точки зрения решения той или иной аналитической задачи	7.1
50	Обработка данных		Процесс изменения данных, приведения их к виду, удобному для использования или соответствующему поставленным требованиям	6.1
51	Озеро данных		Обширное хранилище, в котором собираются необработанные большие данные в исходном формате (без предварительной обработки). Озеро может хранить данные различной структуры	6.2
52	Операционные данные		Буквально: данные о текущих операциях. Данные о клиентах, поставщиках, партнерах и сотрудниках, доступные в процессе онлайн-обработки транзакций и/или полученные из онлайн-базы данных аналитической обработки. Обычно включает транзакционные, контактные данные и общие данные о лицах	2.2

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
53	Открытые данные		Систематизированные данные, организованные в формате, допускающем автоматическую обработку без предварительного изменения человеком, размещенные в интернете в целях неоднократного, свободного и бесплатного использования	3.1.3
54	Очистка данных		Процесс выявления и исправления ошибок и несоответствий в данных с целью улучшения их качества	6.2
55	Платформа данных		Современное хранилище, объединяющее традиционное хранилище данных (DWH) и озеро данных	6.2
56	Подход, основанный на данных	Ориентированный на данные, на основе данных, data-driven	Подход к принятию управленческих решений, в котором решения принимаются не просто исходя из опыта и квалификации или согласно должностной инструкции, а на основе анализа и интерпретации данных должного качества	1.1
57	Полуструктурированные данные		Данные, не имеющие строго определенной структуры, но предполагающие наличие правил, позволяющих выделять отдельные семантические элементы при их интерпретации (прежде всего, правил расстановки тегов и других маркеров, отмечающих и выделяющих элементы данных). Например, файлы, созданные с использованием языка XML, html-страницы	2.2
58	Потоковые данные		Данные, непрерывно передающиеся из разных источников, например постоянно работающих датчиков, видеокамер, микрофонов и т. д. См. также Стриминг	2.2
59	Предиктивная аналитика	Прогнозная, предсказывающая аналитика	Моделирование тенденций и возможных изменений на основе имеющихся на сегодняшний день данных и разработка предположений о том, что может произойти в будущем	9.2
60	Прескриптивная аналитика	Предписывающая аналитика	Комплексная аналитика с целью определения наиболее благоприятного сценария действий и анализ способов реализации этого сценария	9.2
61	Публичные данные		Информация, распространяемая государственными органами, как правило, достоверная и структурированная. Использование публичных данных регулируется законом. В отличие от открытых данных для доступа к ним могут потребоваться определенные действия	2.2
62	Реестровая модель		Модель деятельности, в которой юридически значимыми данными являются записи в реестре. Любые документы в этом случае являются «выписками из реестра» на данный момент времени	3.2.1
63	Руководитель по работе с данными	Chief Data Officer, CDO	Специалист, ответственный за обработку и анализ данных, разработку управленческих решений и использование данных в организации	5.1
64	Руководство данными	Data Governance, DG	Деятельность по осуществлению руководящих и контрольных полномочий (планирования, мониторинга и обеспечения выполнения) в отношении управления информационными активами организации. Одна из функций управления данными, выступает в качестве руководящей по отношению ко всем остальным его функциям	4.4
65	Сбор данных		Процесс перемещения данных в единую область для их последующего использования	6.1
66	Системы управления базами данных	СУБД	ПО, специализирующееся на управлении большими объемами данных и формировании запросов к ним	6.1

№ п/п	Термин	Синонимы и сокращения	Определение	Раздел
67	Стюард данных	Дата-стюард, Data Steward	Сотрудник, обеспечивающий эффективное использование информационных ресурсов и контроль над ними. Важно: в российской практике дата-стюард, как правило, технический писатель, не имеющий широких полномочий	5.1
68	Стратегия управления данными	Data Management Strategy	План обслуживания и повышения качества данных, обеспечения их целостности, регулирования доступа к ним, их защиты и минимизации известных и предполагаемых рисков. См. также Дата-стратегия	4.3
69	Стриминг		Передача данных в потоковом режиме	6.2
70	Структурированные данные		Данные, имеющие строго фиксированную структуру, определяемую формальной моделью данных	2.2
71	Темные данные	Dark Data	Данные, которые собираются, но не используются в обычной деятельности организации	2.2
72	Транзакционные данные		Сведения, отражающие результат изменения данных; относятся к фиксированному моменту времени. Как правило, такие данные порождаются т. н. транзакциями — единицами взаимодействия между разными действующими лицами. Примеры транзакций: покупка, перевод денег, ответ на поступивший запрос и т. д.	2.2
73	Управление данными	Data Management, DM, управление информацией	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	1.3.3
74	Управление информационными активами	Information Asset Management, IAM	Постоянно уточняющийся и расширяющийся набор принципов и концептуальных положений, определяющих подход к управлению данными как стратегически важному активу организации. Его также можно рассматривать как новый, датацентричный образ мышления	1.3.3
75	Управление корпоративной информацией	Enterprise Information Management, EIM	Комплексный подход к управлению данными и информацией на уровне всей организации — как в коммерческих структурах, так и в министерствах, ведомствах и любых больших организациях	1.3.3
76	Утечка данных		Неправомерное раскрытие конфиденциальной информации (конфиденциальных данных), которое может быть умышленным или случайным	8.1
77	«Феодализм данных»		Ситуация, в которой данные представлены локальными информационными ресурсами структурных подразделений (отделов и департаментов), а их интеграция в единые наборы затруднена по административным, техническим и другим причинам	9.2
78	Хранилище данных	ХД, Data Warehouse, DWH	Специальным образом организованный массив данных, обрабатываемый и хранящийся в едином аппаратно-программном комплексе, который обеспечивает быстрый доступ к оперативной и исторической информации, многомерный анализ данных, получение прогнозов, статистики и т. д.	6.2
79	Эксперт по данным в функциональном подразделении	Subject Matter Expert, SME	Эксперт по данным, который работает в своем функциональном подразделении, но при этом может подчиняться еще и CDO; отвечает, в частности, за предоставление стюардам и архитектору информации по терминам, связям, логике формирования показателей	5.1

ПУБЛИКАЦИИ ЦЕНТРА ПОДГОТОВКИ РКЦТ



ГОСУДАРСТВО КАК ПЛАТФОРМА: ЛЮДИ И ТЕХНОЛОГИИ

Практическое пособие для тех, кто участвует в осуществлении цифровой трансформации государственного управления в России.

«Государство как платформа: люди и технологии» развивает идеи доклада «Государство как платформа», выпущенного фондом «Центр стратегических разработок» в мае 2018 года, и содержит конкретные рекомендации, как формировать ИТ-команды цифровой трансформации, а также описание компетенций и технологий, необходимых участникам команды.

www.ranepa.ru/images/News/2019-01/16-01-2019-GovPlatform.pdf



AGILE-ПОДХОД В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ

Методические рекомендации по применению гибких подходов в проектном управлении в органах государственной власти.

Навигатор цифровой трансформации позволяет сформировать единое понимание области применения гибких подходов для управления проектами цифровизации, дает практические рекомендации по использованию Agile при управлении проектами цифровизации в органах исполнительной власти, в том числе с учетом ограничений современной практики госуправления в РФ.

gosagile.cdto.ranepa.ru



МОДЕЛЬ КОМПЕТЕНЦИЙ КОМАНДЫ ЦИФРОВОЙ ТРАНСФОРМАЦИИ В СИСТЕМЕ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Обоснование ключевой роли модели компетенций для формирования успешных цифровых команд.

В детальном описании модели раскрыты ключевые личностные и профессиональные компетенции, описаны их характеристики и поведенческие индикаторы. Показано, почему приоритетом является кадровое обеспечение ЦТ: оно позволит реализовать проекты цифровой трансформации в органах власти на качественно новом уровне.

hr.cdto.ranepa.ru/cm



ОРГАНИЗАЦИОННЫЕ СТРУКТУРЫ И КОМАНДЫ ЦИФРОВОЙ ТРАНСФОРМАЦИИ В СИСТЕМЕ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Результаты изучения организационно-функциональных структур подразделений, ответственных за ЦТ в России, и лучших зарубежных практик.

Подробно анализируются особенности организационных структур в зависимости от их функций, численности и состава персонала, в частности цифровых команд. Представлено описание ключевых ролей системы управления и реализации цифровых проектов.

hr.cdto.ranepa.ru/os_0

Материалы о модели компетенций будут полезны прежде всего руководителям цифровой трансформации, кадровым службам, а также всем категориям госслужащих, специалистам, участвующим в разработке цифровых продуктов и услуг, читателям, интересующимся темой цифровой трансформации в государственном секторе.



САМОИЗОЛЯЦИЯ: РАБОТАЕМ, РУКОВОДИМ, ТРАНСФОРМИРУЕМ

Рекомендации по работе госслужащих в режиме удаленного доступа в условиях противодействия распространению новой коронавирусной инфекции.

В докладе представлены пошаговые алгоритмы, выделены приоритеты, освещен опыт перехода крупных и средних организаций и органов власти на удаленную работу. Собраны рекомендации как руководителей проектных команд, команд разработки продуктов, которые традиционно представляют собой распределенные команды, так и функциональных руководителей ФОИВ, которые уже получили и обобщили опыт перехода на удаленный режим работы.

udalenska.cdto.ranepa.ru

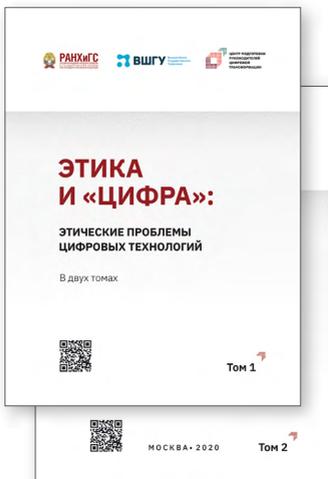


КЛИЕНТОЦЕНТРИЧНЫЙ ПОДХОД В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ

Рекомендации по внедрению клиентоцентричного подхода в организации в виде практико-ориентированного навигатора.

В навигаторе описаны около 50 инструментов изучения клиентского опыта, большинство из которых просты и доступны тем, кто впервые погружается в тему клиентоцентричности. Он отвечает на практические вопросы: какие инструменты подходят для разных проектов? Как обеспечить доступность продукта или услуги? Какие ГОСТы и стандарты нужно знать при создании клиентоцентричного продукта? С чего начать внедрение клиентоцентричного подхода в организации? Каждый может стать клиентоцентричным уже сейчас, освоив в повседневной работе один-два инструмента в качестве первого шага.

cx.cdto.ranepa.ru



ЭТИКА И «ЦИФРА»: ЭТИЧЕСКИЕ ПРОБЛЕМЫ ЦИФРОВЫХ ТЕХНОЛОГИЙ

Взгляд на этические проблемы цифровых технологий преимущественно с точки зрения их применения в государственном управлении.

В первом томе особое внимание уделяется технологиям сбора и обработки цифровых данных, а также искусственному интеллекту и интернету вещей. Во втором томе рассматриваются этические проблемы, возникающие в связи с применением цифровых технологий в различных сферах, где велика роль государства, а также в тех сферах и отраслях, которые могут особенно сильно трансформироваться с переходом в «цифру»: в медицине, образовании, социальных услугах и т. д.

ethics.cdto.center



СТРАТЕГИЯ ЦИФРОВОЙ ТРАНСФОРМАЦИИ: НАПИСАТЬ, ЧТОБЫ ВЫПОЛНИТЬ

Основные принципы создания и применения стратегических документов цифровой трансформации в органах власти и других госорганизациях.

Для государственных и муниципальных служащих доклад может стать удобным и полезным навигатором по разным аспектам разработки стратегии, поможет избежать ошибок, учесть опыт первопроходцев и выбрать свой путь. В книге описаны основные компоненты стратегии цифровизации и ЦТ, исследования, необходимые для ее создания, архитектурный подход к проектированию. На российских примерах показаны необходимые этапы разработки и реализации. Издание адресовано руководителям, кураторам, участникам проектов ЦТ, а также всем, кому интересна эта тема.

strategy.cdto.ranepa.ru



БЕРЕЖЛИВОЕ УПРАВЛЕНИЕ В ГОССЕКТОРЕ. КАК НАЛАДИТЬ ПРОЦЕССЫ

Рекомендации по внедрению современных подходов, основанные на лучших российских практиках в бизнесе и госсекторе.

Навигатор рассказывает о существующих инструментах, кейсах и результатах процессного подхода, плюсах и минусах разных методов и особенностях их внедрения. Адресован в первую очередь руководителям цифровой трансформации (РЦТ) и членам команд, которые занимаются процессной деятельностью. Навигатор будет интересен также сотрудникам процессных офисов (или подразделений, отвечающих за процессное управление в организации), которые хотят больше узнать о цифровой трансформации.

lean.cdto.ranepa.ru



ЭТИКА И «ЦИФРА»: ОТ ПРОБЛЕМ К РЕШЕНИЯМ

Продолжение исследования роли этики в цифровой трансформации.

От описания основных проблем и рисков цифровых технологий в госсекторе авторы второго доклада серии перешли к поиску возможных вариантов разрешения этических дилемм. Особого внимания заслуживает разработанный специально для этого издания фреймворк оценки этичности цифровых решений — инструмент, позволяющий выявить риски и зоны роста информационной системы или другого цифрового решения с точки зрения этики и клиентоцентричности.

И фреймворк, и рекомендации экспертов будут полезны тем, кому предстоит принимать непростые решения в стремительно развивающемся цифровом мире.

ethics.cdto.center/2021

О ЦЕНТРЕ ПОДГОТОВКИ РУКОВОДИТЕЛЕЙ И КОМАНД ЦИФРОВОЙ ТРАНСФОРМАЦИИ

Центр создан в феврале 2019 года под эгидой Минкомсвязи и Минэкономразвития на базе Высшей школы государственного управления РАНХиГС. Финансирование Центра осуществляется в рамках национальной программы «Цифровая экономика».

Директор Центра — **Ксения Андреевна Ткачева**.

Академический директор — **Мария Сергеевна Шклярчук**.

Ключевые направления деятельности Центра

- › Образовательные программы для государственных и муниципальных служащих, ответственных за цифровую трансформацию и развитие российских органов власти. Центр предлагает коммерческим и государственным организациям обучение по темам цифровой трансформации и цифровой экономики, в том числе индивидуальные проекты обучения «под ключ». В 2019–2020 годах отбор на программы Центра прошли более 23 000 госслужащих, в 2021 году — 12 620 человек, в их числе заместители федеральных министров, вице-губернаторы и заместители руководителей федеральных служб, руководители и участники проектных офисов цифрового развития, специалисты муниципальных органов власти. До конца 2024 года более 70 000 человек смогут пройти обучение по программам Центра.
- › Исследовательская работа, разработка аналитических и методических материалов по темам, связанным с цифровой трансформацией госуправления, цифровыми платформами и их экосистемами, реализацией проектов цифровой трансформации, этикой цифровых технологий.
- › Оценка компетенций специалистов и подбор сотрудников в команды цифровой трансформации. Оценкой, подбором, а также ведением Кадрового реестра и портала cdto.work занимается Центр оценки и кадрового развития специалистов в области цифровой трансформации, открытый в мае 2020 года.

О проектах Центра

Одно из важных направлений работы Центра — просветительская деятельность, популяризация цифровой трансформации как в сфере госуправления, так и в других областях. Для этого в Центре разработаны открытые курсы и реализуются спецпроекты:

- › **«Эпоха цифрового развития: основы цифровой трансформации»** — курс на платформе Stepik, созданный в апреле 2020 года на основе образовательных программ Центра и открытый для всех желающих;
- › **«Цифровая трансформация. Быстрый старт»** — второй открытый курс Центра на платформе Stepik, запущен в 2021 году. Темы курса: ИИ, AR/VR, блокчейн, квантовые технологии, робототехника, разработка цифровых сервисов;
- › **База знаний РЦТ** — информационная система для руководителей цифровой трансформации органов власти, где можно найти официальные документы, методические рекомендации, кейсы, а также инструменты для формирования рейтинга РЦТ. Действует с 2020 года;
- › **«Умная нация»** — документальный фильм о цифровой зрелости Сингапура и пути «азиатского тигра» к цифровизации госуправления и общественной жизни;
- › **«Цифровой алмаз»** — фильм о цифровой трансформации Якутии и серия онлайн-встреч с теми, кто ее осуществляет;
- › **«Цифровая весна»** — серия вебинаров о применении цифровых технологий в госуправлении от ведущих экспертов Центра и приглашенных спикеров;
- › **«Циферкаст»** — подкаст Центра о самых актуальных темах цифровой трансформации в госорганизациях и бизнесе;
- › **Аналитический портал «Сумма технологии»** откроется в 2022 году и будет публиковать материалы о цифровизации и цифровой трансформации, реформах в сфере госуправления, описания российских и иностранных кейсов, комментарии и интервью ведущих специалистов по цифровым технологиям, по управлению изменениями и организационной культуре. Отдельный раздел — База знаний — будет ориентирован прежде всего на специалистов из ФОИВ и РОИВ.

УПРАВЛЕНИЕ ДАННЫМИ В ГОССЕКТОРЕ
Навигатор для начинающих

Редколлегия:

Алексей Архипов, Олег Гиацинтов,
Иван Добровольский,
Ксения Киселева, Александр Малахов,
Михаил Петров, Екатерина Потапова,
Виталий Сазонов (научный руководитель),
Николай Скворцов, Мария Шклярчук

Дизайн и верстка: Наталья Балева, Алла Баженова

Корректор: Марк Кантуров

Сайт: summa.technology/reports/data