



# A2 DATA

Управление корпоративными  
данными

# Содержание

- 01 Зачем нужны процессы Data Governance и управление качеством данных
- 02 Как «продать» проект руководству, business case и как вовлечь бизнес-подразделения
- 03 Дорожная карта внедрения и как сформулировать MVP
- 04 Роль централизованной команды Data Quality и взаимодействие с Data Owners и Data Stewards
- 05 Кто такой Data Owner, как его выбрать и назначить
- 06 KPI для команды Data Governance
- 07 Ключевые факторы успеха проекта

01

Зачем нужны процессы Data Governance  
и управление качеством данных

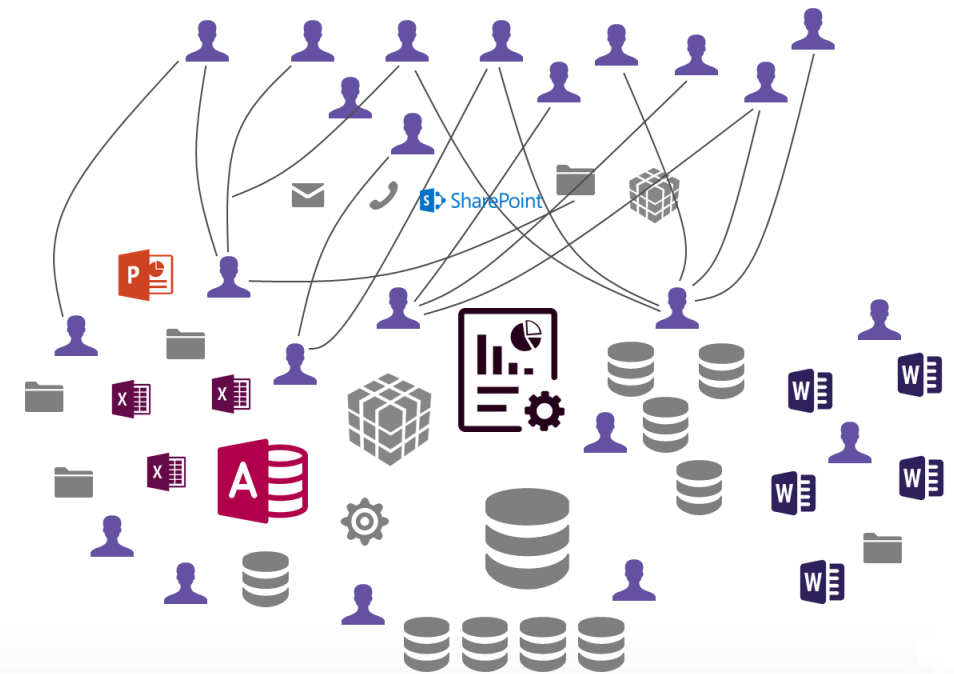
02

Как «продать» проект руководству,  
business case и как вовлечь бизнес-подразделения

# Зачем нужны процессы и решения DG

Крупнейшие организации инвестировали десятки и сотни миллионов долларов в аналитические системы и хранилища данных, но не смогли достичь поставленных целей:

- Невозможность оперативного развития хранилищ данных, слишком длительный time-2-market
- Сложность поиска нужных данных, продвинутые пользователи тратят до 80% времени на поиск данных, вместо анализа данных
- Сложность анализа внедряемых изменений, как изменения в источнике данных повлияет на отчеты и аналитику
- Неясное и неизмеримое качество данных, отсутствие полного доверия к информации из ХД
- Проблемы с качеством клиентских, отсутствие «единого клиента»
- Невозможность сопоставления информации из-за хаоса с нормативно-справочной информацией
- Отсутствие возможности эскалации – нет четко определенной ответственности за данные и качество данных в организации



Потребовался новый подход по управлению информацией – Data Governance

# Ключевые факторы успеха



## Доступность данных



### Скорость поиска

- Корпоративный Google-поиск терминов, метаданных в каталоге данных, владельцев, источников и пр.
- Оптимизация поиска необходимой информации



### Анализ метаданных

- Интерпретация данных – четкое понимание одинаковых и разных по смыслу терминов, атрибутов, расчетных показателей
- Исследование связей между различными объектами метаданных, анализ происхождения, построение data lineage



**УПРАВЛЕНИЕ МЕТАДААННЫМИ**



## Доверие к данным



### Достоверность данных

- Определение качества данных на любом шаге
- Проверка данных при заведении в учетные системы
- Автоматическая очистка, валидация трансформация и стандартизация данных



### Прозрачность состояния данных

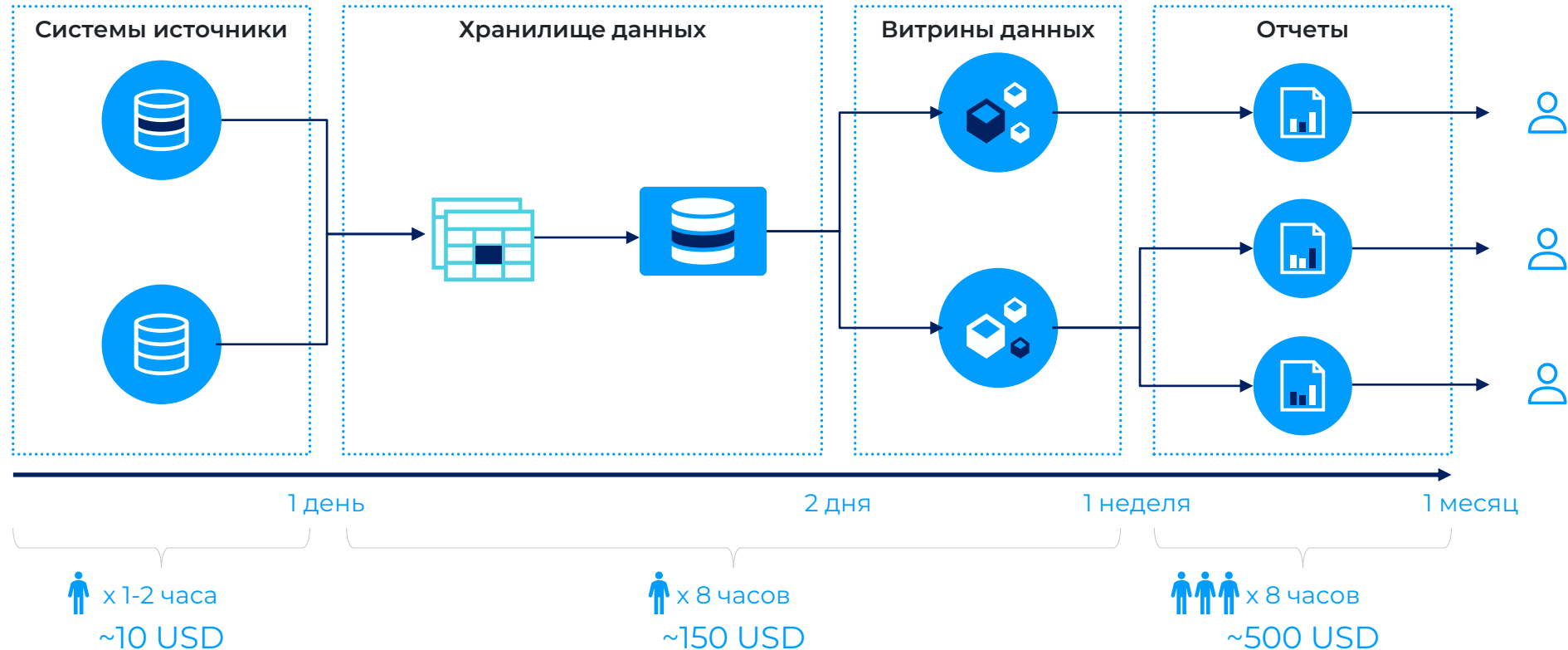
- Профилирование данных – автоматический поиск аномалий, выбросов, ошибочных данных
- Понимание общего состояния данных во всех доступных информационных системах



**УПРАВЛЕНИЕ КАЧЕСТВОМ ДАННЫХ**

# Сценарий: Цена ошибки в данных

Дешевле исправить ошибку на начальном этапе появления данных, т.к. в дальнейшем потребуется затратить больше времени более дорогих специалистов для нахождения причины возникновения и исправления ошибки



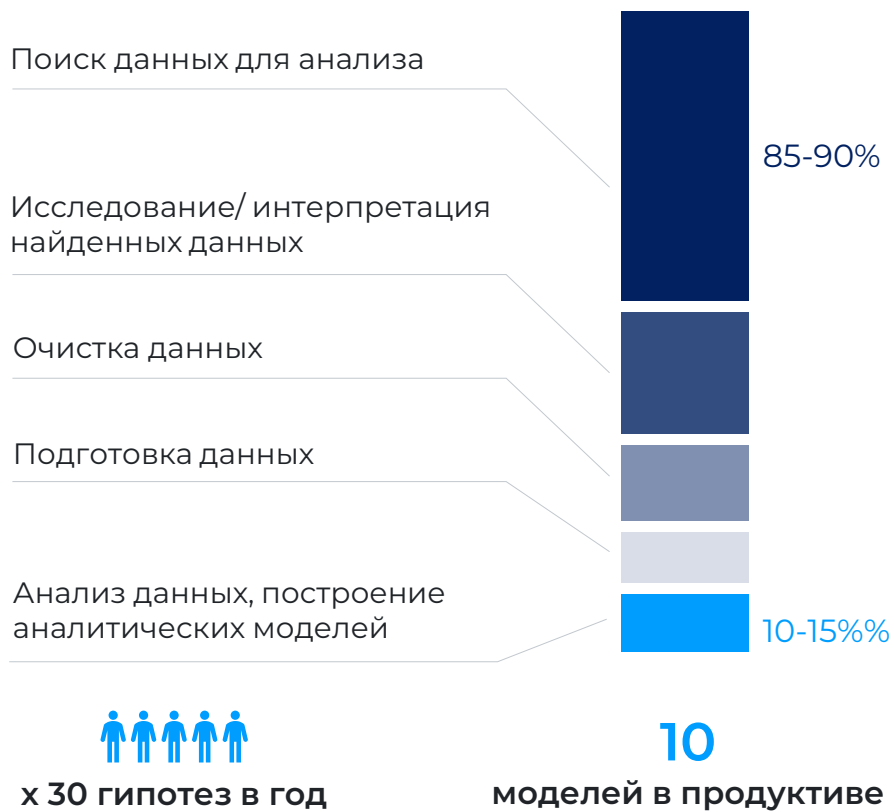
Исправить ошибку в данных на начальном этапе будет стоить несколько часов одного «недорогого» специалиста (оператора данных)

С каждым следующим этапом движения и обработки данных, ошибка из источника распространяется и множится. Найти, на каком этапе возникла ошибка все сложнее. Может потребоваться 1 день ИТ специалиста для нахождения причины возникновения ошибки

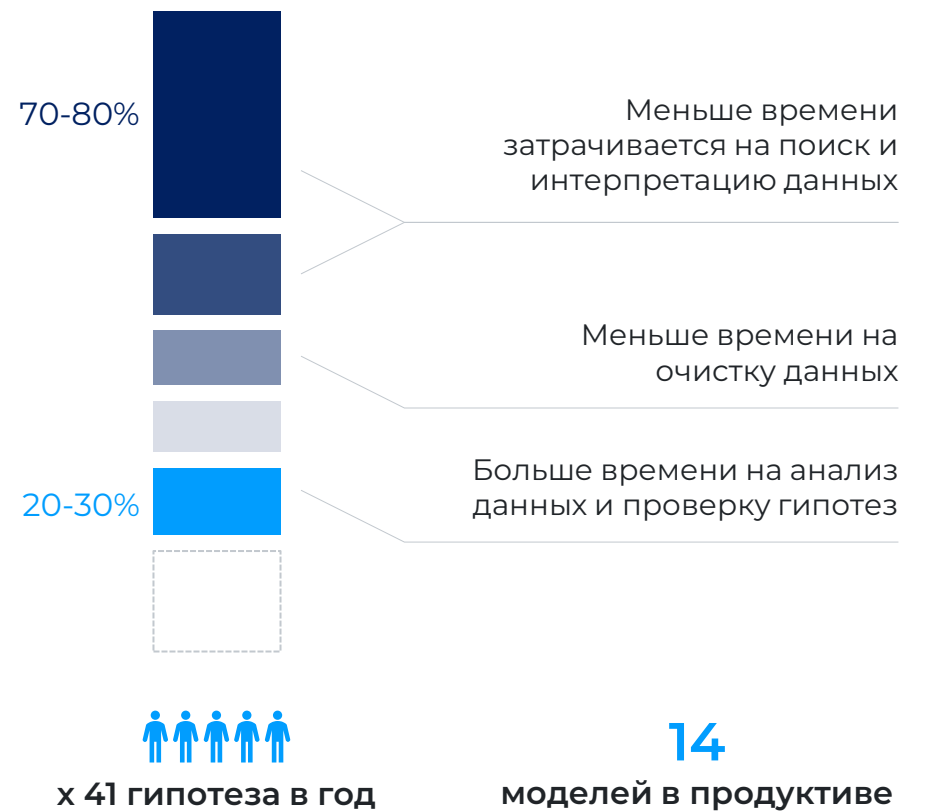
Бизнес-пользователи находят ошибки в данных, сравнивая отчеты разных периодов, и тратят много времени на поиск причины и исправление в отчетах. Одни и те же ошибки, возникшие в разных отчетах, правят бизнес-пользователи разных подразделений

# Сценарий: Доп. прибыль – ML/DS - скорость

Департаменты накапливают очень много данных для применения в своей работе. Data Scientist'ы затрачивают большую часть своего времени на поиск этих данных, их интерпретацию, исследование/ профилирование, выявление особенностей, очистку, объединение, подготовку данных для проведения анализа и моделирования



  
**Рабочее  
время Data  
Scientist**



**+600 000 USD в год**

Условная ценность одной модели 150 000 USD

# Сценарий: Эффективность работы команды аналитиков и использования инфраструктуры

## 01 Эффективность работы команды аналитиков

- Создание контролей качества данных в промышленном инструменте управления качеством данных на 25-30% быстрее, чем реализация контролей вручную
- При анализе данных аналитики и data scientists тратят до 60-80% времени на проверку качества данных
- Сокращение этих затрат повысит эффективность анализа и скорость создания дата-продуктов

## 02 Эффективность использования инфраструктуры

- Сбор и анализ метаданных позволяет выявить паттерны использования данных
- За счет перемещения малоиспользуемых данных в более дешевые системы хранения возможна экономия на стоимости хранения и обслуживания данных.
- Стоимость хранения данных в Hadoop до 10 раз меньше по сравнению с современными MPP - системами

Оценка влияния на Компанию	
Оценка влияния на Компанию при остановки бизнес-процесса	
Через 1-4 часа	
Финансовое влияние	Незначительные потери (50-250 млн. руб.)
Влияние на клиентов компании	Некоторое осложнение в работе с отдельным клиентом
Через 8 - 12 часов	
Финансовое влияние	Низкие потери (250-500 млн. руб.)
Влияние на клиентов компании	Значительные осложнения в работе с малым количеством клиентов
Изменения NPS	Незначительное снижение NPS, до 1-2 п.п.
Через 24 часа - 7 дней	
Финансовое влияние	Средние потери (500-1000 млн. руб.)
Влияние на клиентов компании	Краткосрочное негативное влияние на несколько критичных процессов компании
Изменения NPS	Возможно снижение NPS, до 2-5 п.п.
Через 30 дней	
Финансовое влияние	Высокие потери (1-2 млрд. руб.)
Влияние на клиентов компании	Долгосрочное негативное влияние на несколько критичных процессов компании
Изменения NPS	Снижение NPS, до 2-5 п.п.

Пример отражения информации о финансовом влиянии

## 03 Приоритезация развития и устранения проблем с качеством данных на основании экономической оценки

- Экономическая оценка влияния потери/некачественных данных на бизнес-процессы компании позволяет правильно выстроить приоритеты устранения инцидентов и спланировать развитие, чтобы минимизировать потенциальные потери или увеличить прибыль



# Основные драйверы развития Data Governance



## Аналитика данных внутри Холдинга



### Демократизация данных

- Выстраивание процессов взаимодействия холдинга по эффективной аналитике данных между data-командами
- Качественное описание метаданных для эффективного понимания и использования данных



## Новое КХД Банка



### Эффективность работы с данными

- Описание всех терминов по данным, загружаемым в КХД
- Каталог данных и бизнес-гlossарий – как критичный фактор запуска Self-service BI
- Анализ зависимостей при изменениях



### Качество данных

- Прозрачность в понимании какие проверки качества данных были сделаны разными командами и подразделениями холдинга, повышение доверия к данным
- SLA по инцидентам качества данных
- Снижение затрат на проверки качества данных, осуществляемые разными командами

# 03

Дорожная карта внедрения  
и как сформулировать MVP

# Data Governance MVP. Roadmap развития



## Data Quality

- Определение объектов данных для реализации контролей качества 20-30 пилотных контролей
- Настройка мониторинга и отчетов по реализованным контролям
- Внедрение процесса исправления выявленных проблем качества данных
- Настройка мониторинга

### Принципы выбора контролей для MVP:

- технические проверки
- бизнес-проверки
- наиболее часто встречающиеся ошибки
- наиболее часто используемые данные
- реализация проверок на разных уровнях и системах

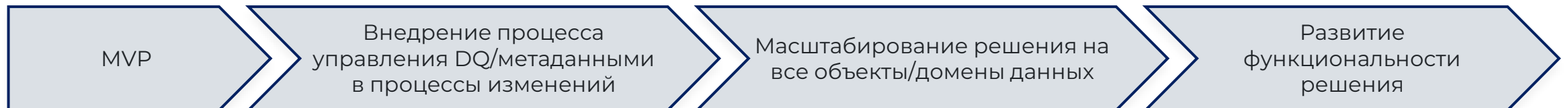


## Metadata management

- Создание модели метаданных
- Описание пилотного домена данных и закрепление владельцев
- Обучение ключевых участников процесса
- Внедрение процесса ведения описания и поддержания актуальности данных
- Настройка мониторинга и контроля

### Ключевые объекты метаданных:

- Глоссарий (термины, показатели)
- Каталог данных (физические объекты хранения)
- Каталог отчетов
- Участники процесса (data owners, data stewards)
- Связи между объектами
- Другое (бизнес-процессы, контроли качества, метрики экономической эффективности данных и т.п.)



# Дорожная карта развития Data Governance



## 01 Data Governance Strategy (Holding)

- Определение общей терминологии, чтобы начать говорить на одном языке
- Определение методологии управления корпоративными данными в структурах Холдинга и дочерних подразделениях
- Требования от Холдинга к дочерним подразделениям по DG
- Определение приоритетных бизнес-кейсов (областей) для начала внедрения
- Определение целевой ролевой модели и подхода к назначению владельцев данных
- Подготовка материалов для обучения сотрудников

## 02 MVP implementation

- Разворачивание платформы Data Governance
- Обучение Data Stewards (централизованная команда)
- Описание терминов по первому бизнес-кейсу – управленческая отчетность
- Профилирование данных
- Определение пилотного набора проверок качества данных
- Внедрение процесса DQ Issue remediation

## 03 Roll-out

- Тиражирование решения на другие предметные области
- Внедрение KPI для ключевых участников процесса
- Tuning решения

# 04

Роль централизованной команды Data Quality  
и взаимодействие с Data Owners и Data Stewards

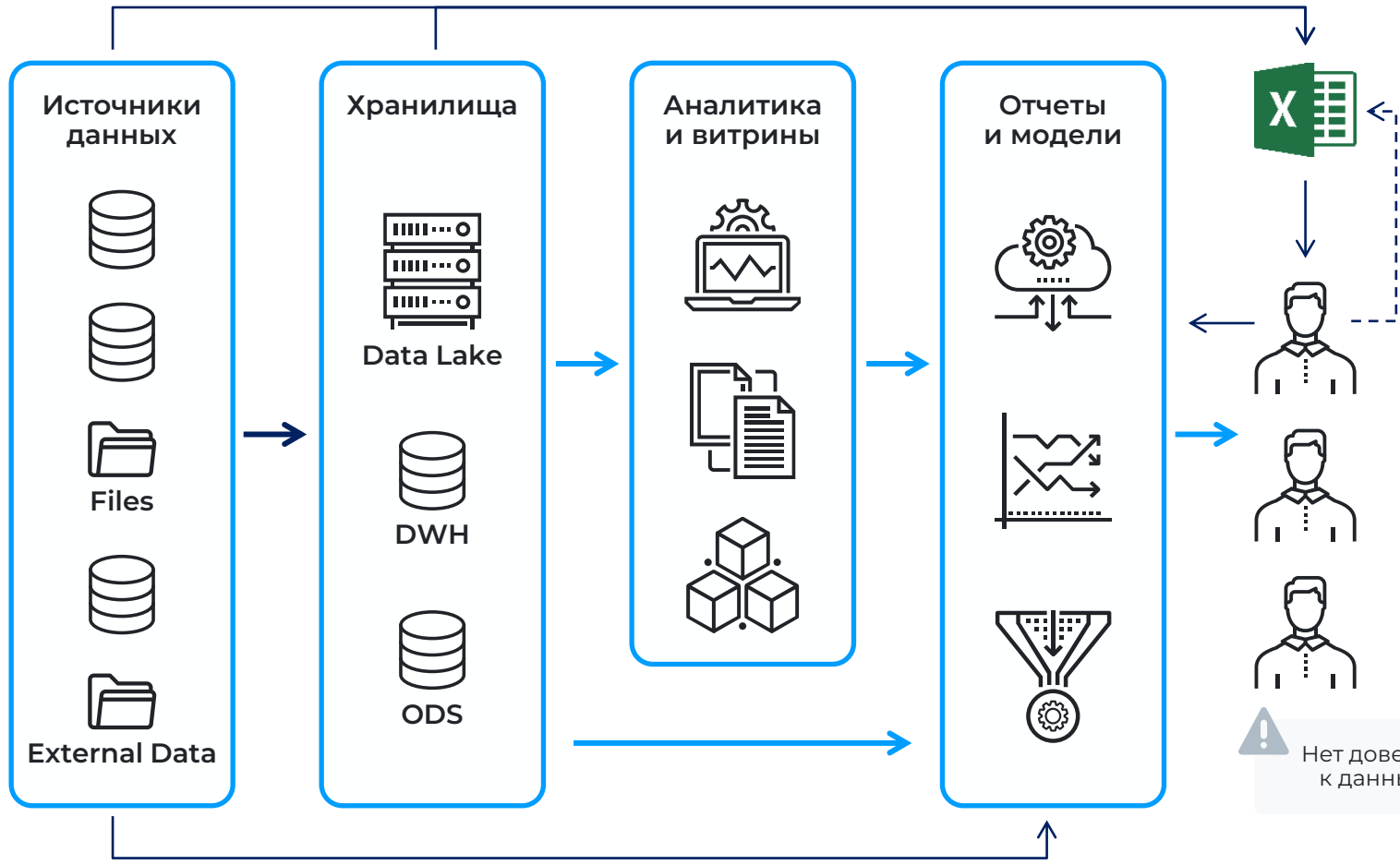
# Введение

⚠ В Хранилищах и витринах, в основном только технические контроли (формат, отсутствие потерь и пр.)

⚠ «Неофициальные» отчеты, собственные отчеты бизнеса, как правило в Excel

⚠ Отсутствие или ограниченные проверки качества данных при вводе

⚠ Не выровненные требования к вводимым данным



⚠ Ручные корректировки данных на финальных этапах (непосредственно в отчетности или при ее формировании):  
**Данные не исправляются в источниках**  
**Многочисленные корректировки одних и тех же ошибок, не редко в авральном режиме**

⚠ Нет доверия к данным

⚠ Отсутствует ответственность за данные (либо определена только на уровне Хранилищ или витрин)

⚠ Необходимость сверки данных в различных источниках (реконсиляция)

# Роль Data Quality Manager

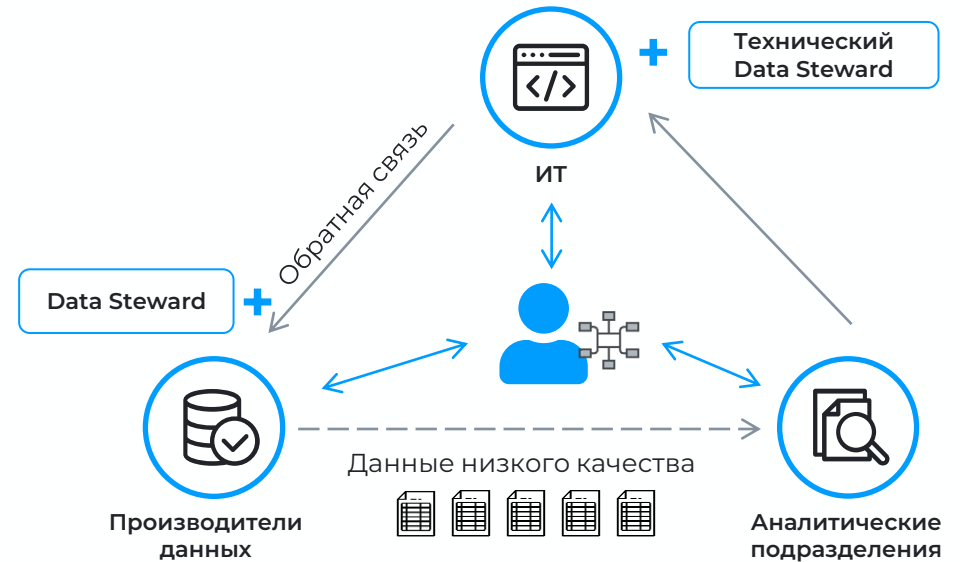
Источником многих проблем является то, что взаимоотношение между производителями данных (владельцами), аналитическими подразделениями и ИТ не выстроено. Ключевой задачей DQ менеджера является выстраивание такого взаимодействия между подразделениями, которое приводит к непрерывному повышению качества данных

## Типичное состояние



Аналитические подразделения, получая данные низкого качества, направляют инциденты качества данных в ИТ. Не являясь ответственными за качество данных в учетных системах, ИТ информируют об этом аналитические подразделения, не перенаправляя жалобы производителям данных. В результате инциденты качества данных не решаются, а данные правятся в итоговых отчетах

## Лучшая практика



DQ менеджер налаживает процесс коммуникации между производителями данных (владельцами), аналитическими подразделениями и ИТ, централизуя обращения по вопросам качества данных и обеспечивая своевременное решение инцидентов качества данных

# Взаимодействие с бизнесом

Наибольшую эффективность процессов управления качеством данных обеспечивает тесное взаимодействие с бизнесом на всех этапах создания дата-продуктов и использования данных

## Управление изменениями

- 01 Сбор требований к качеству данных и реализация контролей на этапе создания нового объекта данных/дата продукта
- 02 Актуализация контролей качества данных в ходе процесса внесения изменений
- 03 Определение требований к качеству данных со стороны системы получателя данных в точках интеграции

Требования к качеству данных формулируются Data Steward. Сформулированные критерии качества являются SLA для команды по управлению качеством данных и основой для оценки уровня качества данных. В случае отсутствия сформулированных требований к качеству данных Владелец данных принимает на себя ответственность за риски использования некачественных данных.

## Управление инцидентами

- 01 Информирование о выявленных ошибках качества данных
- 02 Обеспечение SLA по исправлению ошибок, информирование data stewards о ходе исправления
- 03 Анализ проблем формирование новых контролей качества данных
- 04 Вовлечение data stewards в процессы мониторинга качества данных

Ключевая задача работы с инцидентами – минимизировать ошибки, которые выявляют пользователи данных и максимально информировать о выявленных проблемах качества данных.

## Развитие data-driven культуры

- 01 Коммуникационная платформа для информирования о разных этапах процесса управления качеством данных
- 02 Создание комплекса инструментов для возможности самостоятельного мониторинга и управления качеством данных на стороне пользователей
- 03 Обучение заказчиков корректной постановке требований к качеству данных

Повышение уровня зрелости бизнеса в части вопросов управления качеством данных, повышение уровня ответственности data stewards за данные.



# 05

Кто такой Data Owner, как его выбрать и назначить

# Ключевые роли



## Владельцы данных

### Обязанности

- Отвечает за полноту и качество описания метаданных, обеспечивает выделение необходимых для этого ресурсов
- Принимает стратегические решения по вопросам улучшения качества данных, формирует требования к качеству данных
- Участвует в согласовании изменений, влияющих на состав и качество данных
- Участвует в назначении Data Steward и определяет его временной ресурс

### Мотивация

- КПЭ по полноте и качеству описания метаданных
- КПЭ по качеству данных



## Data Stewards

### Обязанности

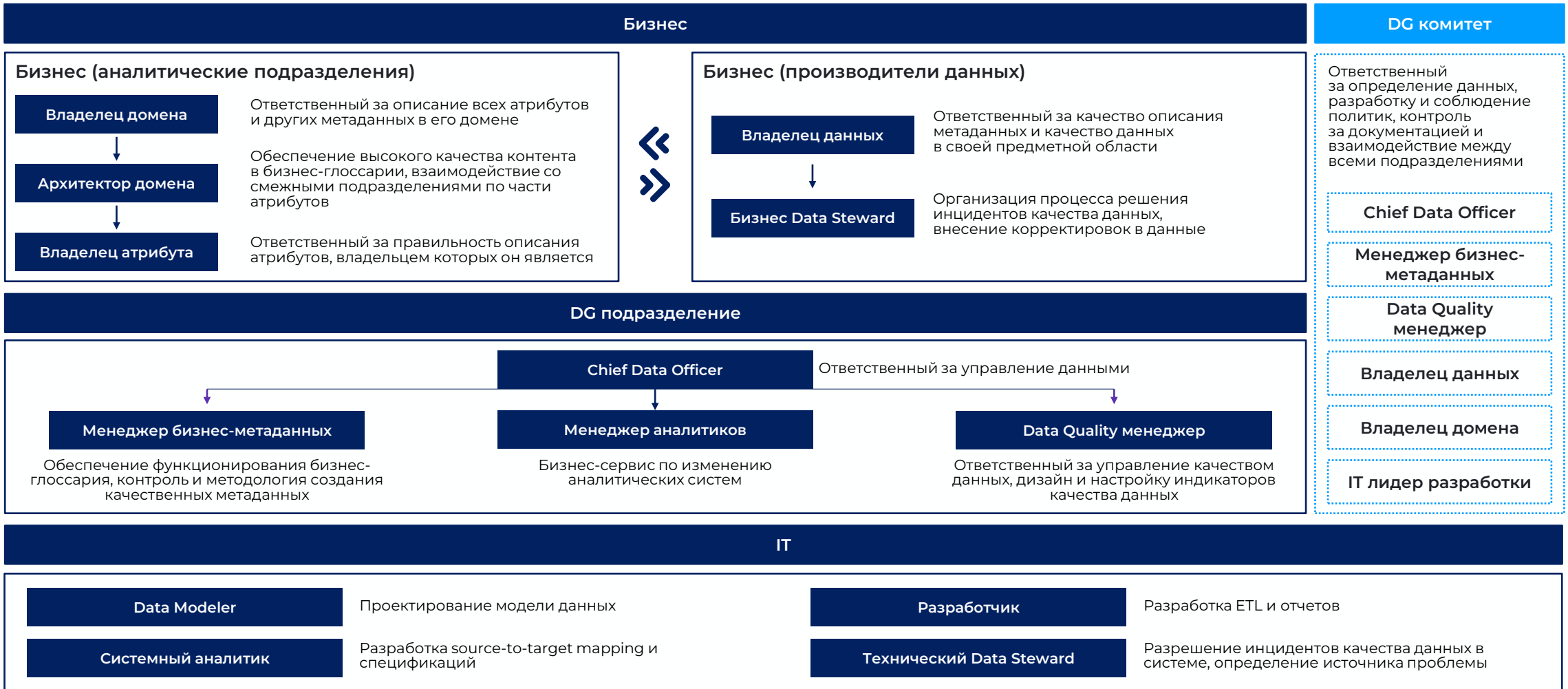
- Обеспечивает полноту и качество описания метаданных
- Анализирует качество данных
- Определяет правила контроля качества данных
- Организует процесс решения инцидентов качества данных

### Мотивация

- КПЭ по полноте и качеству описания метаданных
- SLA и КПЭ по качеству данных

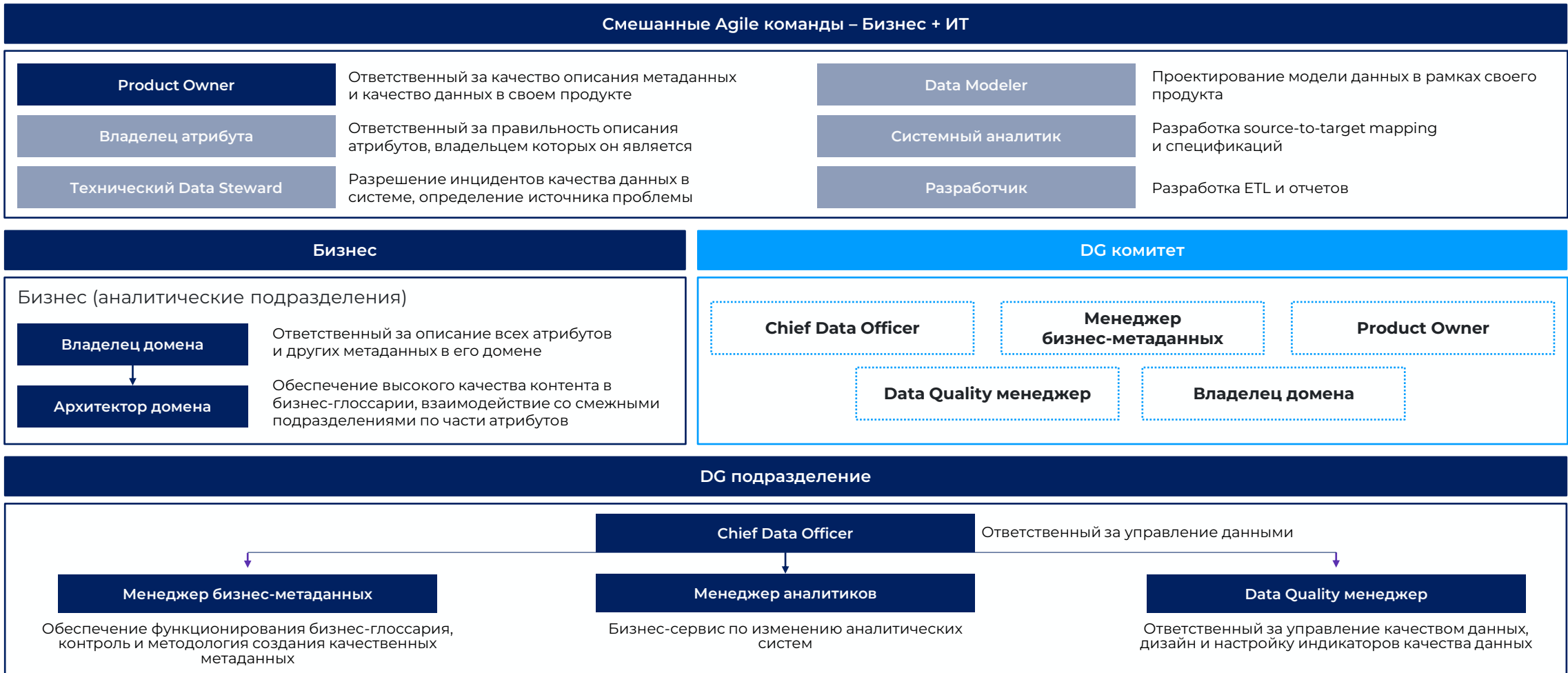
# Управление данными: организационная структура

Эффективное управление данными должно опираться на выстроенные процессы взаимодействия между бизнес-подразделениями и ИТ, а также потребителями и производителями данных. Для этого в организации необходимо предусмотреть новые роли и выполняемые функции



# Управление данными: организационная структура

С распространением методологии гибкой разработки Agile фокус организационной структуры для управления данными смещается на продуктовые Agile команды. Метаданные и Data Quality находятся в ответственности продуктовых команд. Вне Agile команды остаются только органы, контролирующие семантику («общий взгляд на данные») и DG процессы



# Процесс назначения Владельца метаданных домена

Поиск и приоритизация кандидатов на роль владельца метаданных домена осуществляется CDO. Процесс закрепление ответственности за домен производится комитетом по управлению данными



## 2A.3. Data owners should meet prerequisites and at least one of the selection criteria



### Prerequisites

- Sufficient seniority to enact change and deliver needs of data consumers (e.g., director and above)
- Have expertise/ knowledge about the domain (e.g., 3+ years experience in executive role in a domain subject matter)



### Selection criteria<sup>1</sup>

#### 01 Clear business responsibility and power

A stakeholder who has

- Decision making power over structure and creation of domain data
- Financial responsibilities impacted by domain data

#### 02 User

A stakeholder who extracts most of the value from this entity




#### 03 Operation

A stakeholder who manages the team that creates the data

# Each data domain will have a primary cross-functional team managing both daily and strategic activities of the domain

## What are the domain working group roles?

■ IT role ■ BU role

Role	Profile	Data governance responsibilities
 <b>Data Owner</b>	<p><b>Traditional:</b> Director or above  <b>Agile:</b> Chapter lead or Tribe lead</p>	<p>Establish data <b>domain strategy, priorities and goals for data within domain (e.g. next year's main focus will be on "completeness" data quality issues)</b>  <b>Assign and coach data stewards</b> within domain  <b>Have the final say regarding</b> critical decisions in Working Group meetings in topics (such as data element definition, prioritization &amp; security level, retention period, quality issue remediation plan execution)</p>
 <b>Data Steward</b>	<p><b>Traditional:</b> Manager  <b>Agile:</b> Product owner</p>	<p>Be the <b>first point of contact for</b> domain  <b>Create initial drafts for most of the data governance activities</b> (e.g. writing metadata, determining priority and security level, writing data quality rules, identifying <b>data quality issues</b>, creating <b>remediation plans</b>, using <b>data cataloging and quality tools</b>) and <b>present</b> them for <b>Working Group's approval</b>  <b>Provide input</b> on regular standard and <b>process reviews</b> by central data team (SA and PH CDO)</p>
 <b>Data Custodian</b>	<p><b>Traditional:</b> Data analyst  <b>Agile:</b> Team member</p>	<p><b>Support Data Steward</b> in areas that require <b>technical know-how</b> (e.g., <b>write data quality rule scripts</b> on data quality tool, <b>comment on feasibility</b> of data quality rules for the tool, <b>help on understanding feasibility</b> of remediation plans)  <b>Orchestrate efforts</b> required from other IT teams (e.g. <b>invite developers</b> or other business analysts when a <b>more complex script</b> needs to be written/<b>root cause can not be identified</b>)</p>
 <b>Power Data User</b>	<p><b>Traditional:</b> Specialist or manager  <b>Agile:</b> Team member</p>	<p>Attend <b>Working Group</b> meetings and <b>provide input on data elements</b> as well as <b>their needs</b>, concerns and current issues (e.g. provide <b>feedback on prioritization</b>, communicate)</p>





# 06

## КРІ для команды Data Governance



# Data Governance KPI

Employee incentives should be realigned to support behaviors connected to data management best practices. Since enterprise data governance requires cross-functional cooperation, incentives should encourage cross-unit activities and collaboration

	2023 Focus	Long-term Goals & KPIs
 <b>Chief Data Officer</b>	<ul style="list-style-type: none"> <li>• Meeting DWH Roadmap</li> <li>• Implementation of the DQ/DG platform, Data Strategy, policies, procedures</li> <li>• Describe Domains, determine Data Owners</li> </ul>	<ul style="list-style-type: none"> <li>• High level of key users' satisfaction</li> <li>• TCO, infrastructure costs, data availability</li> <li>• DWH adoption rate, quality standards</li> <li>• Data Democratization, Data Monetization</li> </ul>
 <b>Data Owner</b> Executives in Business Units	<ul style="list-style-type: none"> <li>• Domain description in the Data Catalog</li> <li>• DQ Issues Resolution Process implementation for their Domain</li> </ul>	<ul style="list-style-type: none"> <li>• Maintaining a high share of data assets described in the data catalog</li> <li>• Data quality improvement rate</li> <li>• Meeting SLAs for domain DQ incidents</li> </ul>
 <b>Data Steward</b> DG Department/Business Units	<ul style="list-style-type: none"> <li>• Domain description in the Data Catalog</li> </ul>	<ul style="list-style-type: none"> <li>• Maintaining a high share of data assets described in the data catalog</li> <li>• Data Governance training and workshops</li> </ul>
 <b>DQ Team (DQ Manager, Issue Manager, Engineer)</b> DQ Department	<ul style="list-style-type: none"> <li>• Implementation of a DQ platform</li> <li>• DQ Rules with SLAs definition</li> <li>• Implementation of DQ processes and roles for selected data domains</li> <li>• DQ policy is implemented and actively used</li> </ul>	<ul style="list-style-type: none"> <li>• Bank-wide roll-out of DQ processes, roles</li> <li>• Self-service DQ checks set-up capability</li> <li>• DQ improvement rate, SLA coverage</li> <li>• High level of user satisfaction</li> </ul>

# Пример ключевых показателей реализации стратегии Data Governance. Детальные метрики Data Quality

**KPI показывающие эффективность реализации стратегии управления данными. Определяются на основании ключевых целей стратегии Data Governance.**

Цель	Показатель эффективности
Повышение качества данных	Индекс качества критически важных данных Количество инцидентов качества данных Финансовая оценка потерь от использования некачественных данных
Повышение уровня удовлетворенности использования данных	NPS, CSI, T2M
Развитие data-driven культуры	Количество пользователей использующих инструменты self-service анализа данных Количество пользователей прошедших обучение по процессам Data Governance
Повышение эффективности использования данных	% неиспользуемых данных стоимость хранения ITБ данных

## Метрики эффективности процесса управления DQ

- 01 Управление инцидентами качества данных:**
  - Количество обращений (инцидентов) качества данных
  - Количество повторных обращений по ошибкам DQ
  - Контроль SLA закрытия обращений
  - Соотношения выявленных пользователями проблем с качеством данных и проблем, выявленных мониторингом
- 02 Развитие ИТ-системы контроля качества данных:**
  - Размер бэклога доработок в разрезе по типам и командам
  - Статистика по реализации задач в разрезе направлений
  - T2M реализации изменений
- 03 Эффективность системы мониторинга качества данных:**
  - Количество проверок, реализованных в системе
  - Список объектов, с наибольшим количеством ошибок
  - Процент покрытия различных объектов данных проверками DQ
  - Нагрузка на инфраструктуру от проверок DQ
- 04 Развитие уровня зрелости в части качества данных:**
  - Количество требований на реализацию контролей качества данных со стороны бизнес-подразделений
  - Количество пользователей самостоятельно отслеживающих уровень качества данных
  - Количество бизнес-процессов этапы реализации, которых зависят от результатов проверок качества данных

## Основные цели и KPI

### 2023 - начало 2024

- Рост компетенций и увеличение централизованной команды DG & DQ
- Полнота описания данных в новом DWH
- Контроль выполнения SLA по инцидентам качества данных по приоритетным проверкам
- Увеличение числа активных пользователей каталога данных

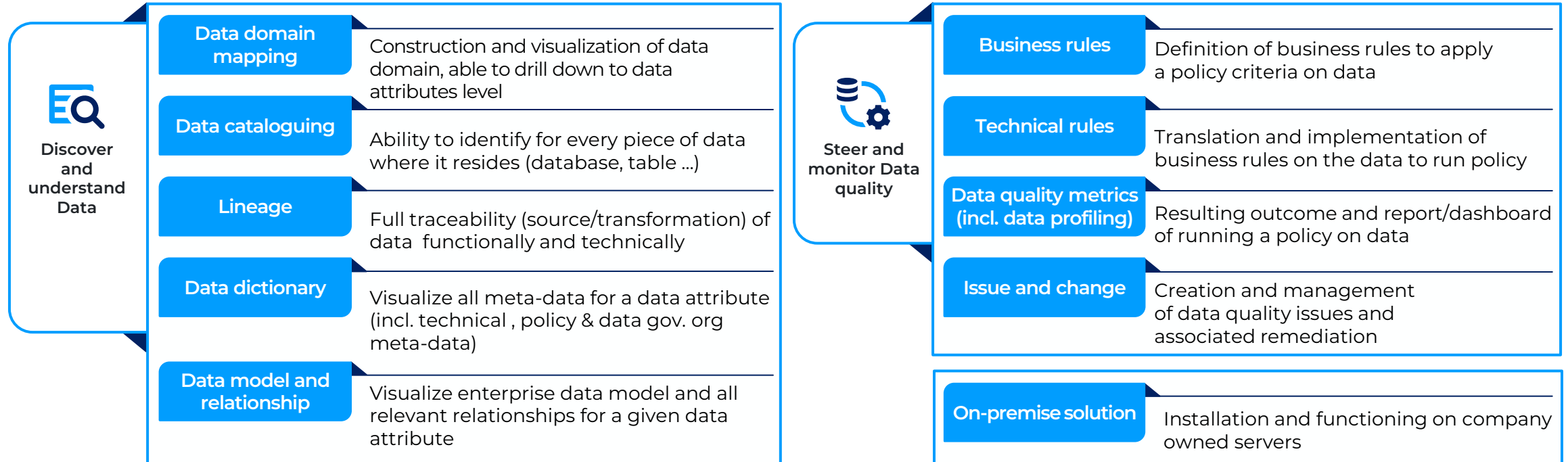
### 2024

- Развитие компетенций в бизнес-командах, переход к гибридной модели работы Data Stewards
- Полнота описания данных по приоритетным бизнес-областям
- Появление KPI по метрикам DQ
- Увеличение числа активных пользователей каталога данных

07

Ключевые факторы успеха проекта

# Ключевые требования и факторы успеха



**Ключевой фактор:** выбирайте не только инструмент, но и команду внедрения, которая имеет проверенный опыт по внедрению инструмента и процессов Data Governance

# Рецепты успеха

## 1 Business first

Внедряемые инструменты Data Governance должны помогать бизнесу и в первую очередь поддерживать цели бизнес-стратегии, а не усложнять процессы

## 2 Внедряем процессы, а не инструменты

В первую очередь изменяем процессы, для встраивания новых подходов управления данными в регулярные бизнес-процессы компании. Новые подходы и инструменты должны упрощать процессы, а не создавать дополнительную нагрузку. Для внедрения нужно выбирать команду с сильной методологией, а не только удобный инструмент

## 3 Data-driven культура

Любые изменения вызывают сопротивление. Поэтому одновременно с началом проектов по внедрению новых подходов и инструментов необходимо начинать программу обучения и популяризации новых процессов

## 4 Коммуникации

Крайне важно разработать и поддерживать адекватную информационно-коммуникационную политику в части процессов управления данными. Проработать получателей и удобные каналы предоставления информации об изменениях, идеях, развитии. Проработать систему информационных панелей и трансляции результатов мониторинга



# A2 DATA



[www.a2data.ai](http://www.a2data.ai)

