



**ТИНЬКОФФ**

# Как не потеряться в больших данных

# Дмитрий Пичугин

Руководитель группы  
управления данными  
Тинькофф



d.pichugin@tinkoff.ru



@Volian

# Проблема поиска данных

К 2025 году в мире будут собраны рекордные 175 ZB данных. Данных в компаниях становится настолько много, что в них сложно ориентироваться даже специалистам. При этом принимать решения нужно все быстрее

## Влияние

Согласно исследованиям среднестатистический аналитик тратит на поиск данных до 50% своего рабочего времени

## Критичность



Задержки при принятии быстрых и качественных решений серьезно вредят бизнесу в долгосрочной перспективе

# Масштаб проблемы

5000 +

аналитиков

25+

Кластеров СУБД

150 000 +

Таблиц

9000 +

Отчетов BI

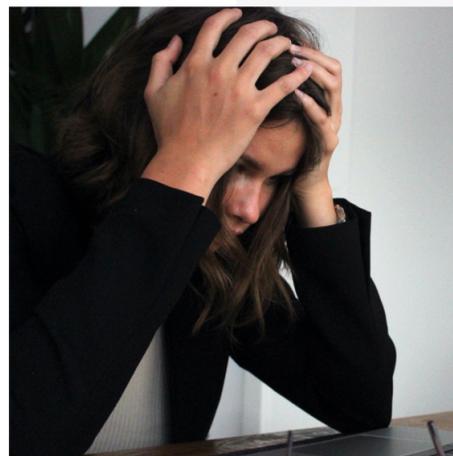
38 000 +

Self-Service  
Zeppelin notes

14 000 +

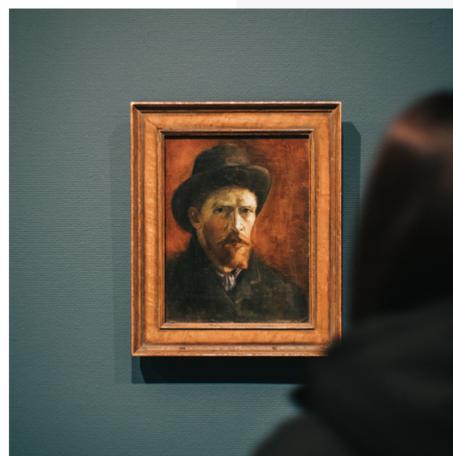
ETL pipelines

# Как оценивали поиск данных наши пользователи



40%

Испытывали затруднения  
в поиске данных



«Поиск метаданных в этом  
банке – это искусство, ему  
нельзя научить»



«Хранилище в текущем виде  
непознаваемо для  
пользователей»

Проблема поиска  
данных — драйвер  
развития  
управления  
метаданными  
в Тинькофф



# Управление метаданными

Планирование, организация и контроль деятельности по обеспечению доступа к качественным и интегрированным метаданным

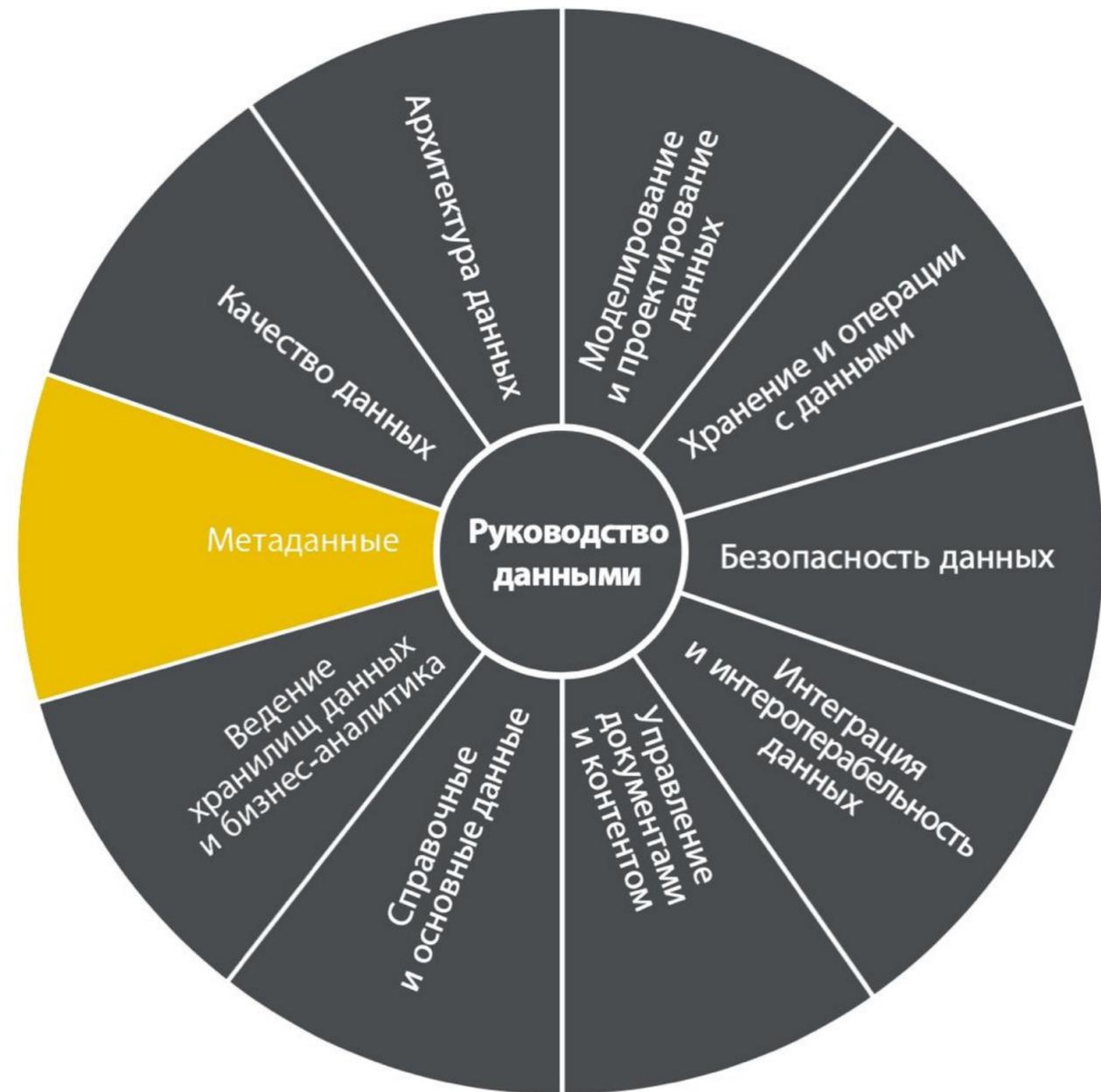
## Решаемые проблемы

Поиск данных

Базис для других активностей

## Инструменты

Каталог данных



**DAMA-DMBOK2 Data Management Framework**

Copyright © 2017 by DAMA International

# Метаданные

01

Данные о данных

02

Включают в себя широкий спектр сведений

03

Одни и та же информация в разных ситуациях может играть роль как данных, так и метаданных

Пример: описание сериала, жанр, список актеров

# Пути решения проблемы

 Предоставление пользователю доступа к метаданным

Обеспечение доступности метаданных и удобного поиска по ним

 Повышение качества и полноты метаданных

Предоставление пользователю полных и качественных метаданных



**Предоставление пользователю  
доступа к метаданным**

# Каталог данных



## Каталог

Единая точка правды



## Поиск

Инструмент для обнаружения  
нужных данных



## Data lineage

Движение данных от источника  
по точкам применения и обработки

# Решения на рынке



Open  
Metadata



DataHub



Alation



Amundsen



Apache Atlas



Collibra



+1 от Тинькофф

# Data Detective

Кейс Тинькофф

# Значительно сократили время на поиск данных

## Задача

Аналитики в среднем тратят до 50% рабочего времени на поиск данных

Задача — сократить время на поиск и сделать его не сложнее поиска товара на маркетплейсе

## Решение

Старались сделать платформу такой, чтобы ей могли пользоваться аналитики любого уровня

Чтобы достичь целей, мы построили платформу на основе принципов:

- Удобство пользователя
- Гибкость продукта

## Результат

Создали платформу, которая экономит время аналитиков на поиск данных — они могут потратить его на помощь в принятии важных для бизнеса решений

Основные функции:

- Единое хранилище данных компании
- Простые инструменты Data Discovery

# Как выглядит Data Detective

The screenshot displays the Data Detective interface. At the top left, there is a navigation link "Вернуться назад". Below it is a search bar containing the text "claim". The main area shows a list of reports, each with a "Логический отчет" label, a title, a breadcrumb "DATA PLATFORM > LOGICAL REPORTS", and a description. A tooltip is visible over the first report, "IssDRep", providing additional details. On the right side, there is a filter sidebar with a "Скрыть фильтры" button. The sidebar includes a section "Искать совпадения по" with radio buttons for "Только по имени", "Искать везде" (selected), and "Тэги". Below this is a "Типы" section with checkboxes for various report types, including "Логический отчет" which is checked. At the bottom of the sidebar, there is a "Очистить фильтры" button.

< Вернуться назад

claim

Логический отчет **IssDRep**  
DATA PLATFORM > LOGICAL REPORTS

Отчет показывает количество выпущенных эмбоссированных пластиковых карт в день + процент от предыдущего. Отчет формирует информацию за месяц.

Логический отчет **Циклы ATM**  
DATA PLATFORM > LOGICAL REPORTS

ATM.

Логический отчет **SmeFmQRep**  
DATA PLATFORM > LOGICAL REPORTS

Отчет показывает основные показатели по эффективности работы SME в пресечении фрода и...

Логический отчет **DispSRRep**  
DATA PLATFORM > LOGICAL REPORTS

Показывает обращения в ДПСиОП и среднее время их обработки

Логический отчет **DisplssRep**  
DATA PLATFORM > LOGICAL REPORTS

Количество претензий по эмиссии, поступающих в ДПСиОП.

Логический отчет **InvClaimSRRep**  
DATA PLATFORM > LOGICAL REPORTS

Клиенты и информация из SR с суммой более 50 000 в рублёвом эквиваленте, поступившие вчера.

Логический отчет **c2c-documents**  
DATA PLATFORM > LOGICAL REPORTS

Транзакции по которым поступил диспут

✖ Скрыть фильтры

Искать совпадения по

Только по имени

Искать везде

Тэги

Типы

Атрибут

Колонка

Cut

База данных

SDP: Домен

SDP: Сущность

Spheradian: Фича

Guillotine

Job

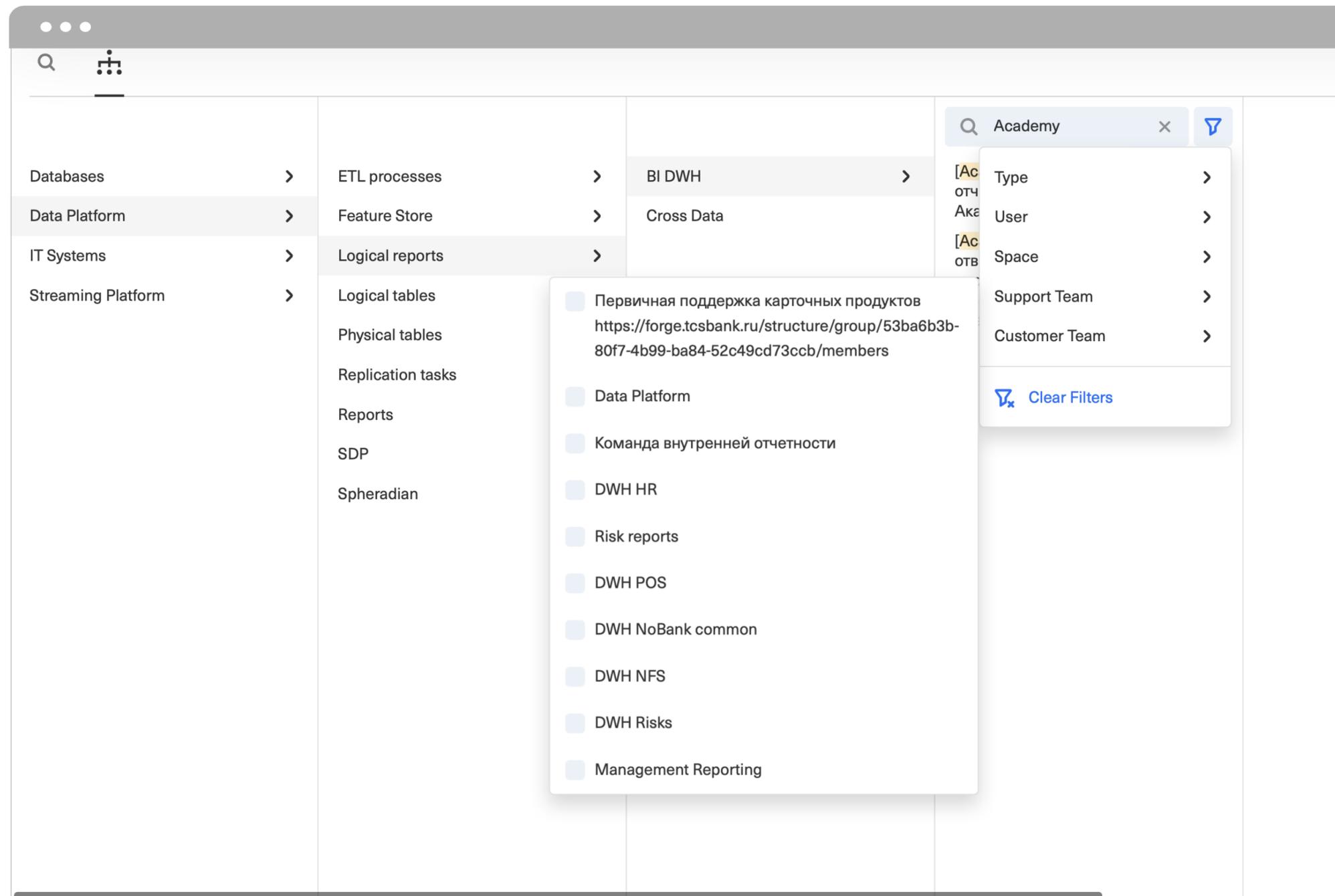
Kafka Topic

Логическая колонка

Логический отчет

Очистить фильтры

# Как выглядит Data Detective



# Как выглядит Data Detective

The screenshot displays the Data Detective interface for the table `prod_dds.abm_point`. The interface is divided into a left sidebar and a main workspace.

**Left Sidebar:**

- Search bar with a magnifying glass icon.
- Menu icon (three horizontal lines) highlighted with a yellow border.
- Link labeled "Свернуть" (Collapse).
- Section header "Свернуть список" (Collapse list).
- Tree view showing the hierarchy: `prod_datavault` > `prod_dds` > `abm_point`.
- Under `abm_point`, a list of fields is shown: `abm_point_rk`, `abm_point_type_dk`, `point_nm`, `product_cd`, `abm_resource_dk`, `abm_activity_dk`, `abm_process_dk`, `abm_driver_dk`, `processed_dttm`, `abm_channel_dk`, and `abm_point_id`.

**Main Workspace:**

- Header: `prod_dds.abm_point` Таблица (Table).
- Breadcrumbs: DATA PLATFORM > PHYSICAL MODEL > GREENPLUM > PROD\_DDS.
- View toggles: Card and Lineage (selected).
- Lineage diagram showing data flow from source tables to target tables through various load processes. The source table `prod_dds.abm_point` is expanded to show its fields. The target tables include `bi_v_dds.abm_point`, `prod_usermart.abm_chain_downhill`, `prod_emarti.abm_chain_allocation`, `prod_usermart.abm_chain Uphill`, `prod_marti.abm_point_x_cost_unit_m`, `prod_emarti.abm_chain`, and `prod_emarti.abm_chain_alloc_g_point_type_act`.
- Intermediate load processes shown include: `USERMART LOAD ABM CHAIN DOWNHILL`, `EMARTI LOAD ABM CHAIN ALLOCATION`, `USERMART LOAD ABM CHAIN UPHILL`, `MARTI LOAD ABM POINT X COST UNIT M`, and `EMARTI LOAD ABM CHAIN`.
- Bottom right corner contains search and zoom icons.



Подробнее тут!



Под капотом  
каталога данных



Как мы создавали  
каталог данных





# Повышение качества и полноты метаданных

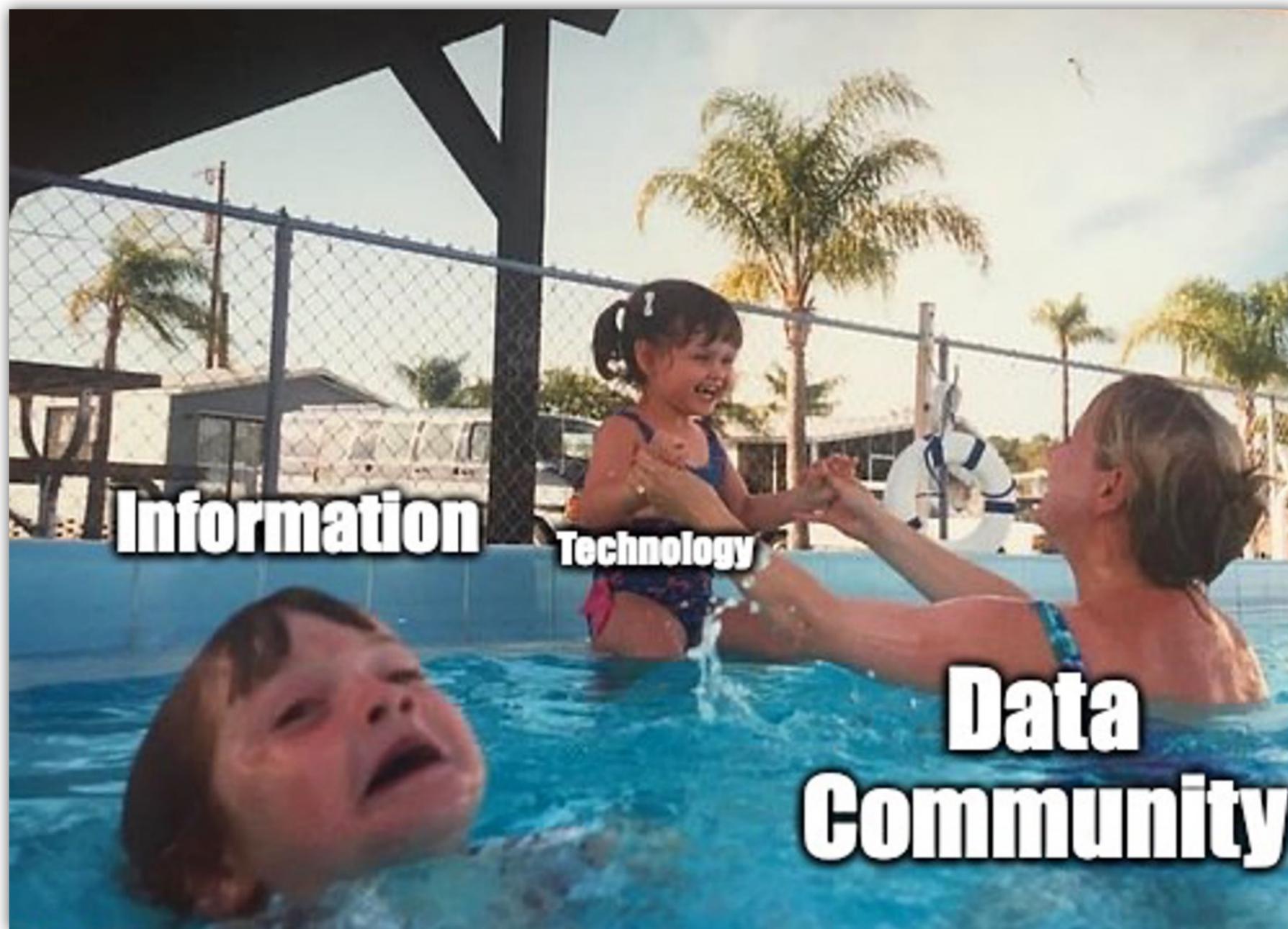
... и вот здесь все становится сложнее

Так исторически сложилось

**I**nformation

**T**echnology

Так исторически сложилось



# Выделенная команда руководства данными



## Проектирование

Создает и поддерживает процессы управления метаданными



## Внедрение

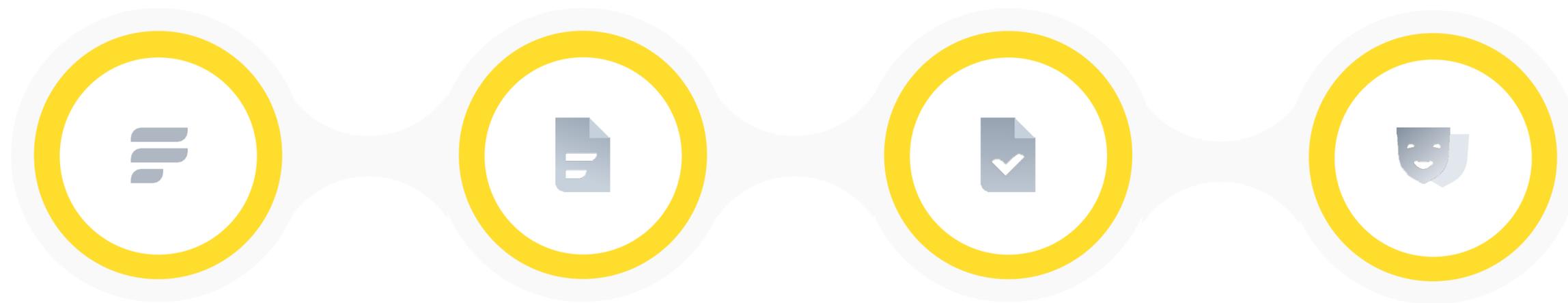
Помогает внедрить процессы внутри команд Тинькофф. Обучает и разрешает спорные ситуации



## Контроль

Мониторит метрики выполнения процессов и оказывает помощь командам

# Фреймворк работы с метаданными



Стандарт ведения

Стандарты и правила  
ведения  
метаданных

Инструмент  
ведения

Инструмент,  
обеспечивающий  
удобный ввод

Мониторинг

Проверки  
корректности  
ведения

Ролевая модель

Описание участников  
процесса и их областей  
ответственности

Люди и устоявшиеся  
в умах процессы —  
основные трудности  
на пути к внедрению  
руководства  
данными



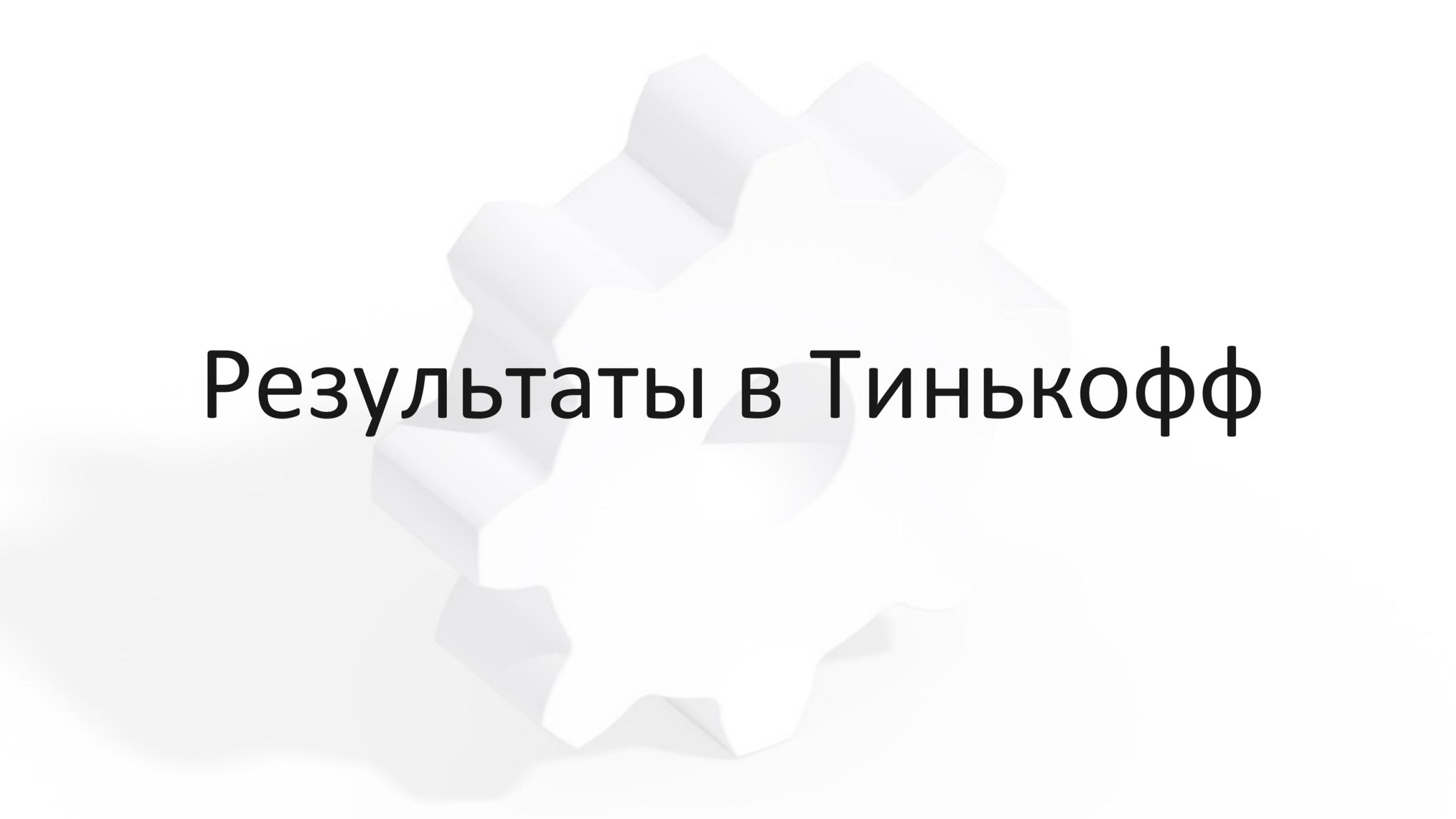
**Подробнее о  
команде**

Data Governance  
в Тинькофф



Как мы строим  
Metadata  
Management





# Результаты в Тинькофф

# Основные цифры

30+

Систем в едином  
репозитории

1200+

Еженедельных  
пользователей

85%

Пользователей  
мигрировали с wiki

54%

запросов на поиск  
от бизнеса

7,9

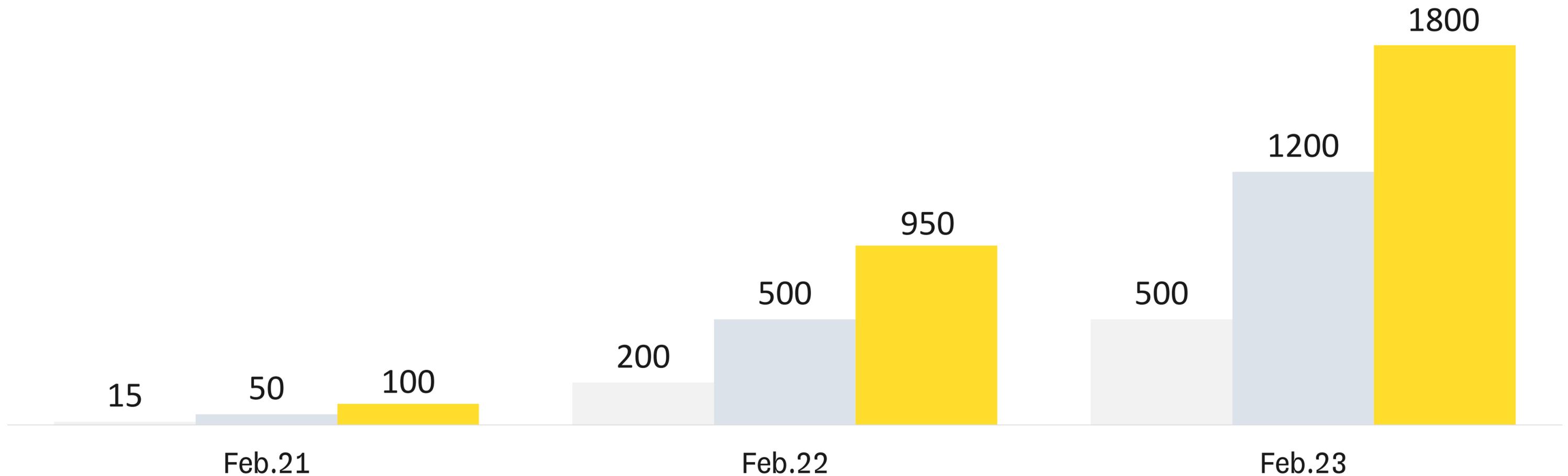
NPS аналитиков

20+

Дополнительных  
сценариев  
использования

# Динамика роста пользовательской базы Data Detective

MAU WAU DAU



# Результаты команды руководства данными

10 000+

Описаний важных объектов

87%

Среднее качество описаний

40%

Важных объектов имеют владельцев

1

бизнес-гlossарий

7,9

NPS аналитиков

# Повторим важные тезисы

- 01 Данные — необходимый компонент успеха компании
- 02 Данных стало так много, что аналитики тратят большую часть своего времени на поиск необходимой информации
- 03 Каталог данных — это инструмент для поиска и централизованного хранения метаданных, то есть контекстной информации для понимания данных
- 04 Функция управления метаданными помогает получать контекст для каталога данных
- 05 Помните, что на перестройку процесса поведения людей нужно время