

MAGNIT DATA

Внедрение платформы управления данными



...про доверие к данным

Олег Молчанов

Руководитель направления по метаданным и
качеству данных



О чём рассказ?

1. Данные Магнита в цифрах
2. Предпосылки проекта
3. Дорожная карта
4. Инструменты
5. Кейсы: как было, как стало?
6. Ожидаемые эффекты
7. Q&A

Зачем?

Предпосылки проекта внедрения платформы
управления данными

Данные Магнита в цифрах

ПЛАТФОРМА ДАННЫХ

>1 ПБ

Данных

10 тыс

Пользователей отчетов и ad-hoc аналитики

200+

Систем источников

КОРПОРАТИВНОЕ ХРАНИЛИЩЕ ДАННЫХ

420+ Тб

Данных

2,2 тыс

Пользователей ежемесячно

>5 тыс

Витрин данных

159 млн

Среднее количество запросов в месяц

80+

Бизнесовых песочниц

DATA-СЕРВИСЫ (запросов в месяц)

100+ на информацию

200+ на изменение и обслуживание

160+ на доступ

2022: предпосылки проекта



В Магните **много разнородных данных**: возникают проблемы с поиском, пониманием, переиспользованием и изменением.



Отсутствует **понимание уровня качества данных** по ключевым активам: работа с качеством ведется реактивно

О проекте

Цели и задачи проекта на старте

- ▶ Развернуть и сконфигурировать **инструменты** в двух средах
- ▶ Разработать **концепцию управления данным** с использованием инструментов
- ▶ Описать ~ 100 терминов в **бизнес-гlossарии**
- ▶ Разработать 25 **проверок качества данных**
- ▶ Отсканировать 6 систем-источников с помощью **Каталога данных**
- ▶ Спроектировать **интеграции** с системами Jira и 1С СУМД (IdM)
- ▶ **Мигрировать** существующие наработки в части бизнес-гlossария
- ▶ **Перенять экспертизу по сопровождению и развитию инструментов**

Дорожная карта

Ноябрь '22

Декабрь '22

Январь

Февраль

Март

Апрель

Май

Июнь

Июль

Август

Сентябрь

Этап 5. Подготовка и проведение ОПЭ по DQ, сканирование дополнительных систем-источников

Настроено 40 проверок качества данных

Получена инфраструктура для Прод контура

Этап 4. Настройка решения DQ, ОПЭ всех сканирований

Инструменты развернуты в Прод

Мигрированы существующие наработки глоссариев

Этап 3. Проектирование решения DQ, сканирование 6 систем-источников (ETL, BI, DWh)

Получена инфраструктура опытно-промышленного контура

Этап 1. MVP (глоссарий + каталог для КХД)

Этап 2. ОПЭ (Глоссарий + Каталог для КХД)

- Развернуты все инструменты на тесте
- Заведены 84 термина глоссария
- Согласовано ТЗ по управлению данными
- Проведено демо на РГ по результатам MVP



Этап 6. Кастомное сканирование Магрепорта

03.05.2023

Инструменты

*«Платформа является движущей силой внедрения
Политики и Стратегии по управлению данными в
Магните»*

Кратко об инструментах

The screenshot shows the Magnit Data dashboard interface. At the top, there is a navigation bar with the Magnit Data logo and menu items: "Управление данными", "Поиск", "Карты", "Мои элементы", and "Создать". On the right side of the navigation bar, there is a search icon, a notification bell with a "1", a user profile for "Олег Молчанов", and a help icon. Below the navigation bar, there are browser tabs for "Главная", "Описание предме...", and "Моя панель монит...".

The main content area features a large heading: "Добро пожаловать в Magnit Data". Below this, a subtitle reads: "Корпоративная платформа для всех, кто работает с данными и аналитикой". A vertical list of navigation links is displayed on the left side, each with an icon and a right-pointing arrow:

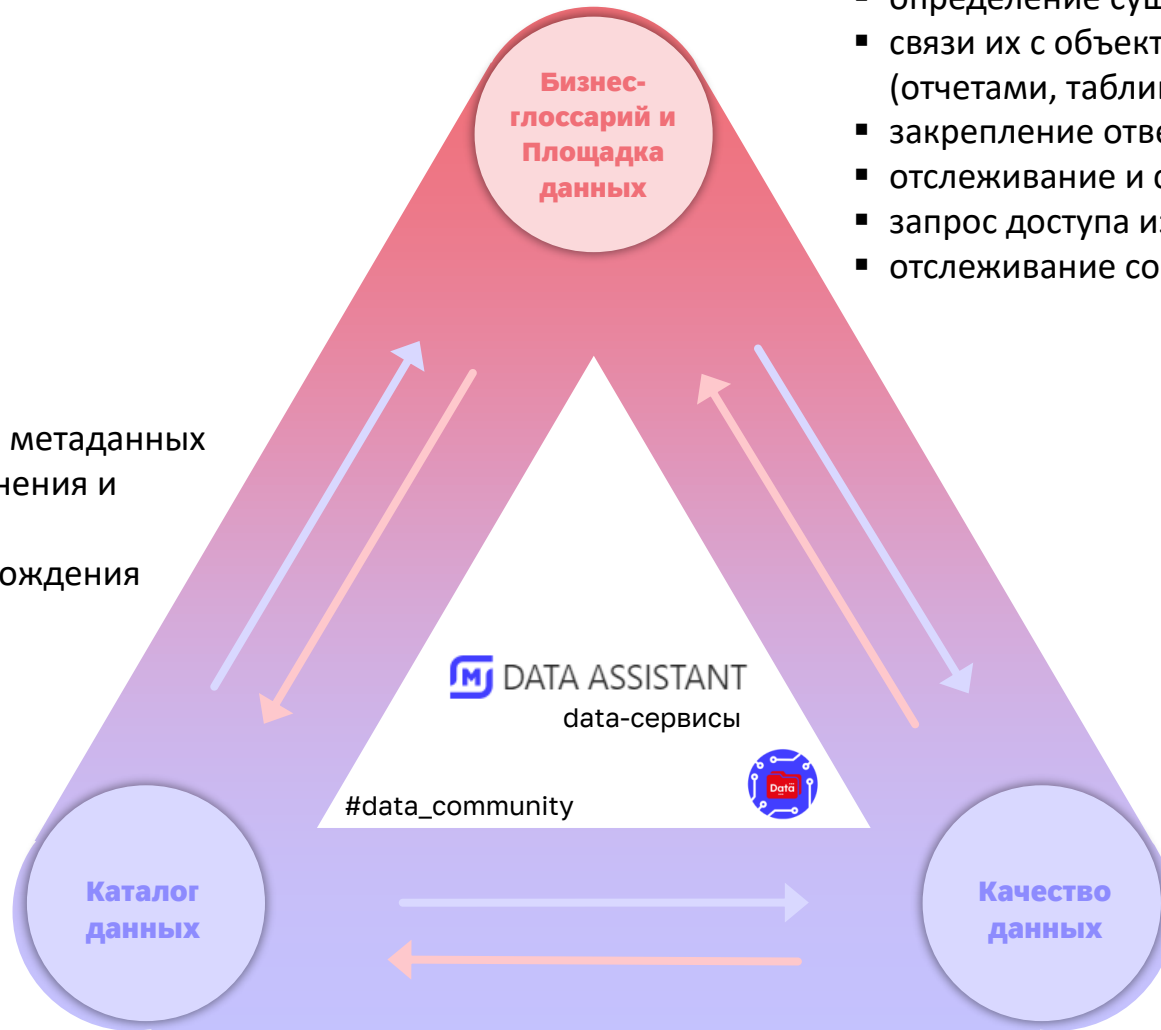
- Новому пользователю →
- Процессы управления данными →
- Бизнес-гlossарий →
- Каталог данных →
- Data Community →
- Площадка данных (beta) →

On the right side of the dashboard, there is a large graphic consisting of overlapping semi-circles in shades of orange and red. Overlaid on the right side of the dashboard is a chat window titled "DATA ASSISTANT". The chat window has a search icon, the title "DATA ASSISTANT", and a close button. The chat content includes:

- A greeting: "Привет, Олег, я помогу тебе сориентироваться в данных и аналитических отчетах Компании."
- A timestamp: "Оператор 10.10.2023 9:37"
- A prompt: "Выбери интересующий раздел в интерактивном меню:"
- A list of interactive buttons:
 - Узнать о Корпоративной платформе данных
 - Получить доступ к данным
 - Задать вопрос; сообщить об ошибке в данных КХД
 - Найти данные в КХД
 - Заказать новые данные в КХД
 - Тест метода /chat_with_gpt/
- A second timestamp: "Оператор 10.10.2023 9:37"
- A text input field at the bottom: "Введите сообщение..." with icons for emojis, thumbs up, and a copy icon.

Кратко об инструментах

- автоматизированный сбор метаданных из систем интеграции, хранения и визуализации данных;
- построение графов происхождения данных (data lineage);
- профилирование данных.



- определение сущностей, показателей и аналитик;
- связи их с объектами физической реализации (отчетами, таблицами);
- закрепление ответственности за данные;
- отслеживание и согласование изменений;
- запрос доступа из единого окна;
- отслеживание состояния качества данных.

- постановка на расписание проверок качества данных;
- автоматическое информирование об отклонениях;
- измерение качества данных.

Главная Поиск Данные, к которым есть доступ (tbd q4) Мои заказы (tbd q4) Данные, которыми я владею

Площадка данных

Добро пожаловать на Площадку Данных Magnit Data! Здесь ты сможешь найти информацию по главным продуктивным витринам КХД и основе аналитической отчетности Компании.

Витрины

В данной категории представлены промышленные витрины корпоративного хранилища данных (КХД). Для удобства витрины сгруппированы по подкатегориям в соответствии с предметными областями.

Отчеты

В данной категории представлены промышленные аналитические отчеты. Для удобства отчеты сгруппированы по подкатегориям в соответствии с бизнес-направлениями.

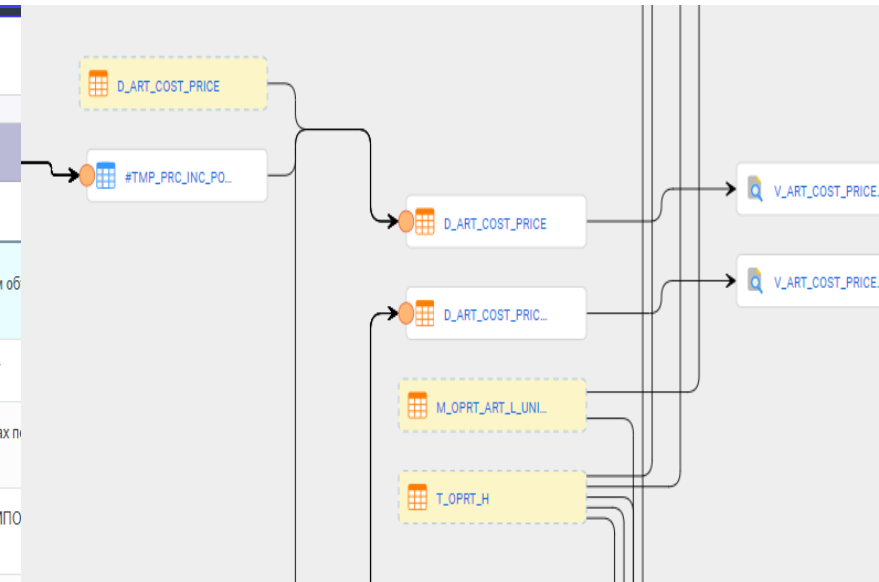
ГЛОССАРИЙ 48 из 2881

НАБОРЫ ДАННЫХ 210 из 5760

КАЧЕСТВО ДАННЫХ 2 из 96

Список Панель мониторинга

Имя	Тип	Определение
Заказы	СУЩНОСТЬ	Включает в себя показатели, связанные с заказанным об
Настройки автозаказа	СУЩНОСТЬ	Показатели, участвующие в алгоритме Автозаказа ТТ
РасхСр	ЭЛЕМЕНТАРНЫЙ	Расчитанный/установленный расход средней в штуках п [Развернуть]
Тип МПО	АНАЛИТИКА	Номер МПО, где: 1 - ОПВ; 2 - МПО шт. из набора; 3 - МПО [Развернуть]



Площадка данных

Описано 5700+ промышленных отчетов и витрин

Интеграции:



Бизнес-гlossарий

- Концептуальная модель (области, сущности, показатели, аналитики)
- Паспорта витрин КХД, отчетов и справочников МДМ
- Правила качества данных
- Реестр информационных систем
- Карта бизнес-процессов
- Организационная структура

Каталог данных

11 подключенных систем-источников

7,1 млн загруженных объектов данных (метаданные)

Внедрили, и?

Было

Разные инструменты, развиваемые децентрализованно разными командами

Частично разобранные взаимосвязи объектов данных.
Ряд источников данных – черный ящик
Нет единого бизнес-гlossария

Разные процессы и ответственные для разных информационных активов (отчеты, витрины, справочники)

Разрозненные процессы и инструменты согласования изменений

Не возможности подписаться на изменения

Стало



НАЙТИ ДАННЫЕ

Единое общедоступное окно поиска информационных активов. Business lineage от показателя в отчете до поля в системе источника



ПОНЯТЬ ИХ ТОЧНОЕ ЗНАЧЕНИЕ

Data lineage с простым подключением новых источников
Общекорпоративный инструмент для ведения бизнес-гlossария и выравнивания понятийного поля в рамках всей компании



ЗАПРОСИТЬ К НИМ ДОСТУП

Возможность запросить доступ к активу (отчет, витрина) из единого интерфейса



ИНИЦИИРОВАТЬ, СОГЛАСОВАТЬ И ОТСЛЕЖИВАТЬ ИЗМЕНЕНИЯ

Единый инструмент с автоматическим workflow
Возможность задать вопрос в разрезе карточки объекта
Возможность отслеживать изменения через почтовые уведомления

Было

Стало



ИЗМЕРИТЬ КАЧЕСТВО ДАННЫХ

Отсутствует инструмент. Не измеряется.

Дашборды с качеством данных в разрезе активов



ПОВЫСИТЬ КАЧЕСТВО ДАННЫХ

Отсутствуют процессы по управлению качеством данных

Автоматизированный процесс взаимодействия с бизнесом
Прозрачная ответственность за активы



СЛЕДИТЬ ЗА КАЧЕСТВОМ ДАННЫХ

Отсутствует регулярный мониторинг, о проблемах узнаем от бизнеса

Информирование об отклонениях
Общедоступная визуализация текущего состояния и динамики качества

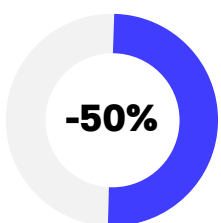
Цель и ожидаемые эффекты

Повысить эффективность компании за счет:

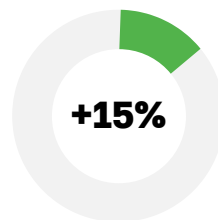
- обеспечения прозрачности информационных активов и знаний о них;
- повышения доверия к существующим наработкам и создания условий их переиспользования;
- разделения ответственности за данные и работу над повышением их качества

В 2024 году планируется разработать и измерить количественные метрики для оценки влияния на Компанию внедряемых инструментов.

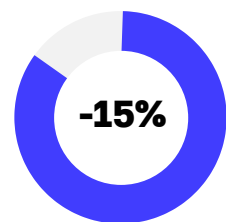
Оценочно ориентируемся* на следующие показатели:



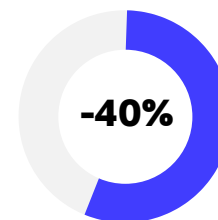
Быстрее поиск данных *
времени на поиск
необходимого показателя и
получение доступа к нему



Выше доверие к данным
к использованию
аналитики для принятия
решений на основе данных



Быстрее разработка
time-to-market при
разработке новых витрин и
отчетов



Быстрее анализ
времени на разбор
происхождения данных и
импакт анализ



Закругляемся

Ценные советы



СДЕЛАЙТЕ MVP СВОИМИ РУКАМИ



НЕ ВНЕДРЯЙТЕ ИНСТРУМЕНТЫ В ОТРЫВЕ ОТ ПРОЦЕССОВ



ВЫБИРАЙТЕ РЕШЕНИЯ, ГДЕ «КАЧЕСТВО ДАННЫХ» И «ГЛОССАРИЙ» ИНТЕГРИРОВАНЫ



ИСПОЛЬЗУЙТЕ ОПЫТ КОМПАНИЙ И ПРОВЕРЕННУЮ МЕТОДОЛОГИЮ

Читай нас на:

HABR



[Habr.com/ru/
company/magnit](https://habr.com/ru/company/magnit)

VC



[Vc.ru/magnit-tech](https://vc.ru/magnit-tech)

Magnit



www.magnit.tech

Q&A

Молчанов Олег

Руководитель направления по метаданным и
качеству данных

molchanov_oe@magnit.ru

8 (995) 110 08 12

