

DATA GOVERNANCE

ЧТО ЭТО, ЗАЧЕМ, И С ЧЕГО НАЧАТЬ

Андрей Вихров, «Сеть Связной»

Обо мне

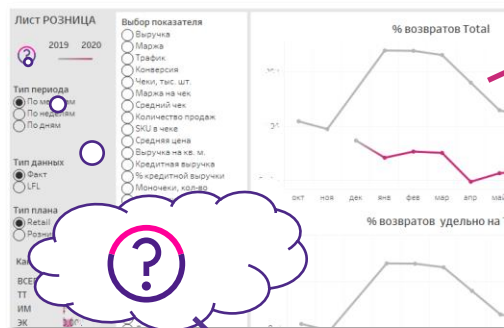
- Создавал и развивал аналитические системы в крупных компаниях, в качестве бизнес-архитектора и владельца систем
- Для эффективного выполнения этих функций давно начал применять элементы DG, поначалу не зная, что они так называются)
- Получил международные сертификаты в области управленческого и бухгалтерского учета (CMA, ACCA)
- В настоящее время отвечаю в «Связном» за инициативы по внедрению Data governance и системы дэшбордов для топ-менеджмента

О «Связном»

- «Сеть Связной» – крупнейшая розничная сеть в сегменте высоких технологий
- Более четырех тысяч магазинов в 1200 городах России
- Интернет-магазин (топ-10 интернет-магазинов), Apple premium reseller CStore
- Более 23 тысяч продавцов-консультантов, которые обслуживают более миллиона покупателей в день
- Ключевой фокус работы Связного – стать проводником в мир технологий для самой широкой аудитории
- Помимо гаджетов компания продает различные услуги, включая дополнительные гарантии, страховки и кредиты на устройства

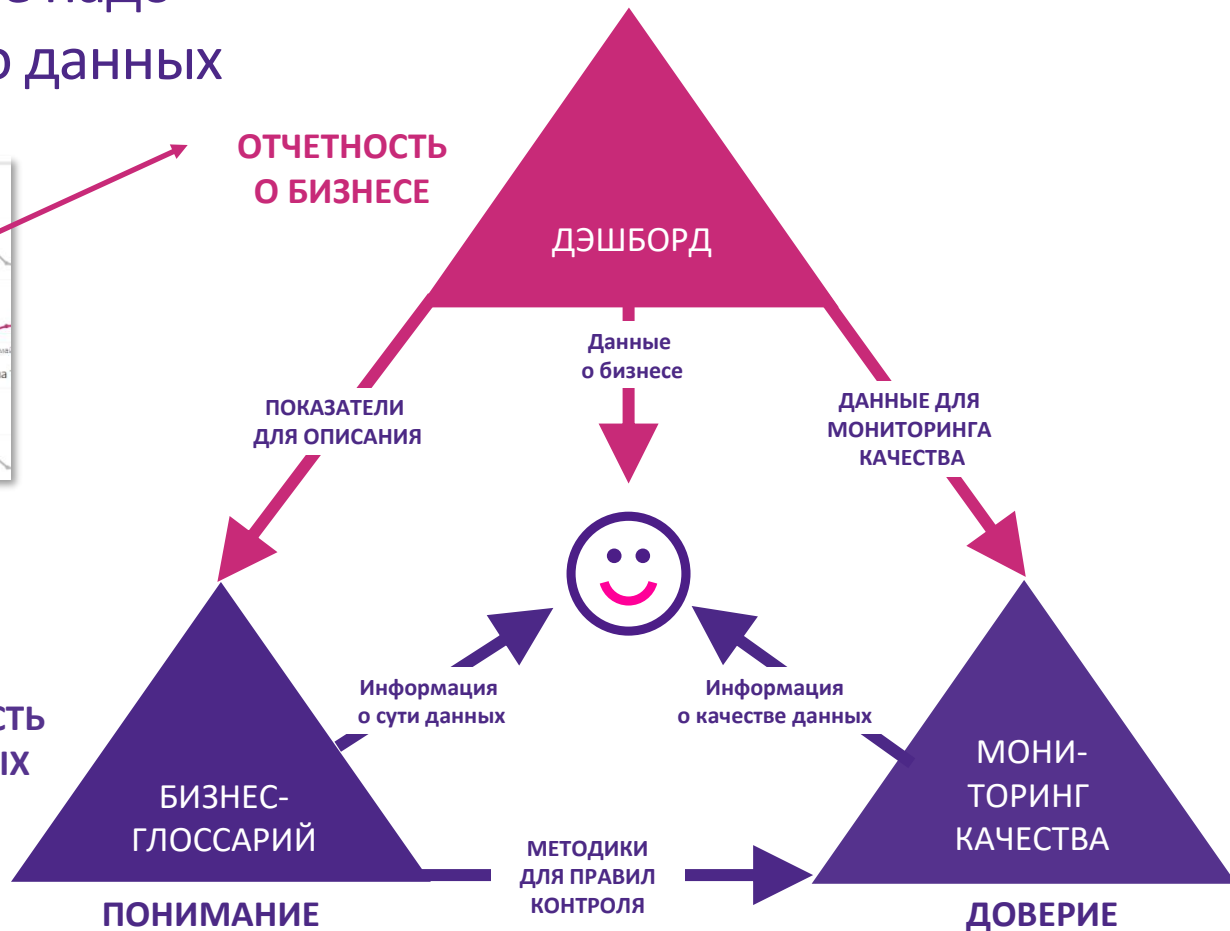
DATA GOVERNANCE

К отчетности о бизнесе надо добавить отчетность о данных



ОТЧЕТНОСТЬ
О ДАННЫХ

ОТЧЕТНОСТЬ
О БИЗНЕСЕ



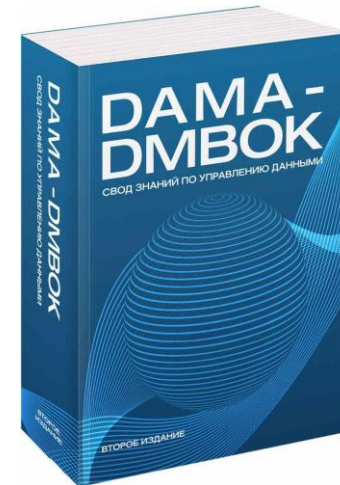
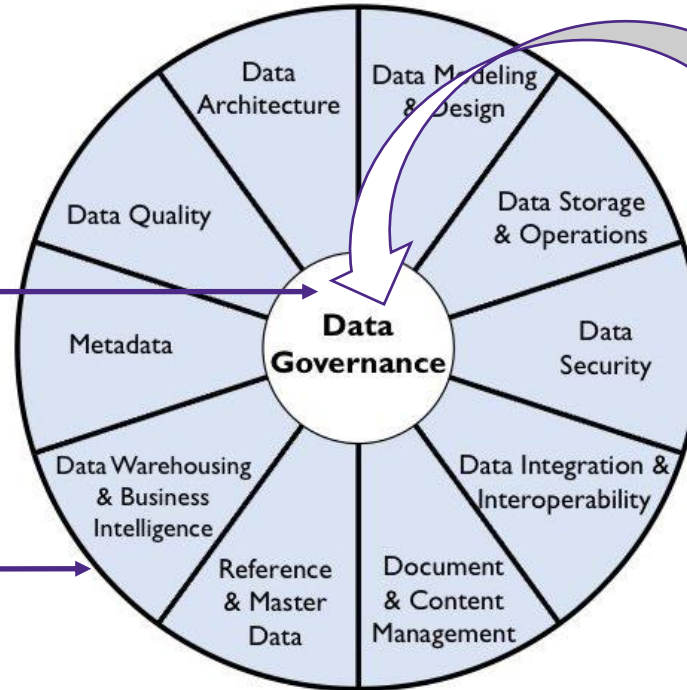
Data governance – связующее звено бизнеса с дата-менеджерами и ядро data management

Data governance

- определяет цели, «что»
- doing the right things
- бизнес бэкграунд

Data management

- достигает цели, «как»
- doing things right
- технический бэкграунд



Copyright© 2017 DAMA International

Данные

Информация, получаемая из систем
(объект управления!)

Качество

Степень соответствия данных
заданным критериям качества

Управление

Процессы управления:

- (plan) - формирование стандартов
- (do) - работа по стандартам
- (check) - мониторинг отклонений от стандартов
- (act) - корректирующие действия

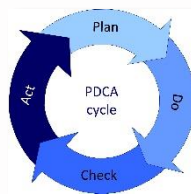
Организация

- Определение субъектов (дата стюарды, комитеты)
- Организация их участия в управлении качеством

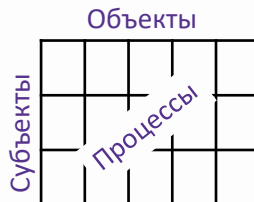
Отчетность по данным:
бизнес-гlossарий

Отчетность по качеству:
мониторинг качества
данных

Цикл Деминга



Матрица ролей

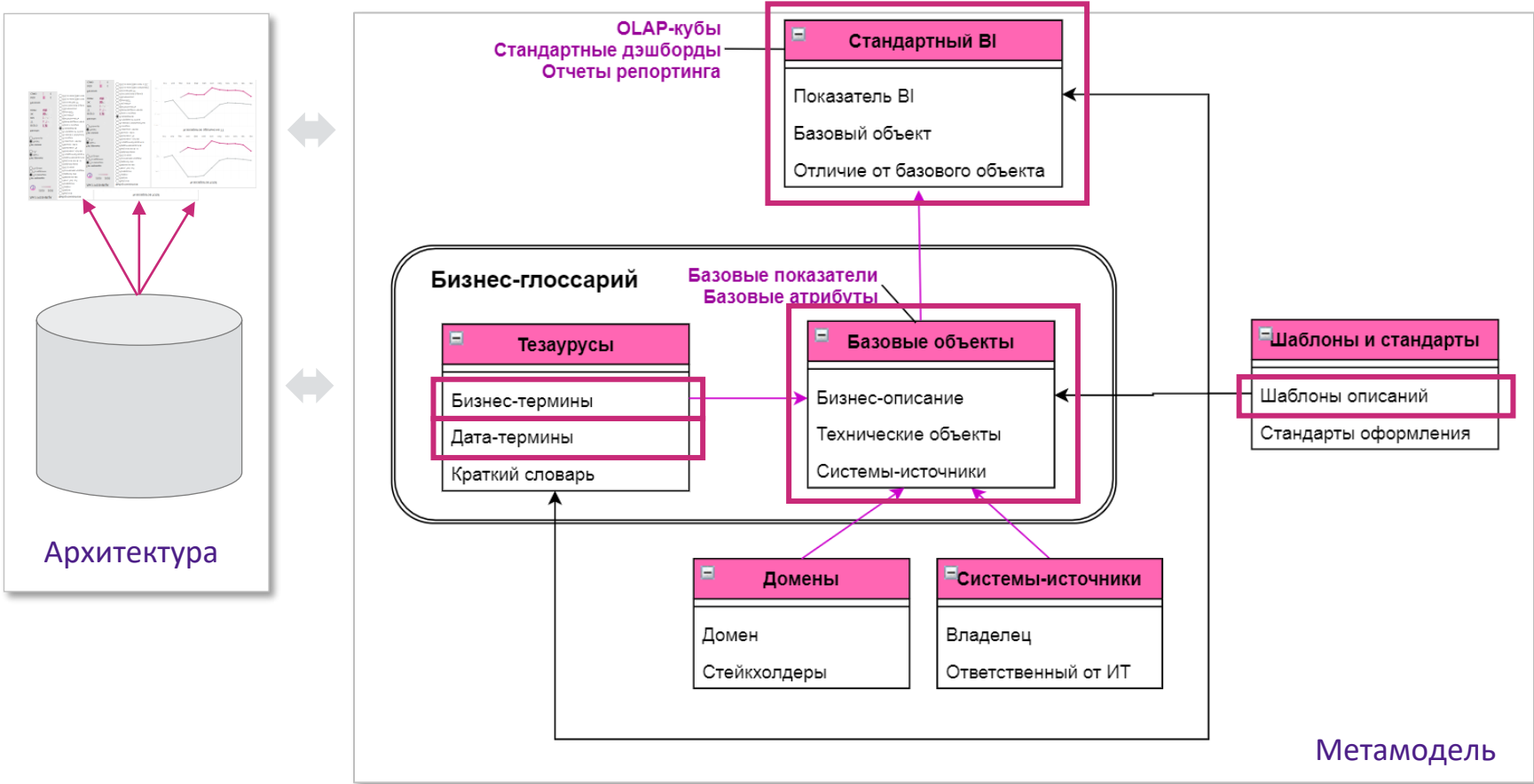


Data Governance -

Организация Управления Качеством Данных

БИЗНЕС-ГЛОССАРИЙ

Метамодель - модель данных для бизнес-гlossария



Описание BI (дэшборда) – точка входа для пользователя

| № | Показатели инфоблока | Базовый показатель | Краткое определение базового показателя | Различие / расчет |
|---|----------------------|--------------------|---|---|
| 1 | Выручка с НДС | Выручка | Сумма денежных средств, причитающихся к получению компанией за продажу продуктов (товаров, услуг или платежей). Включает НДС. Измеряется в руб. | По категории "Кредиты" = "Маржа плюс скидка полная" + "Вознаграждение за кредит". По группе категорий "Нетоварные" также включает 'Маркетинг услуг'. |

Базовый показатель

Создал(а) Викров Андрей Александрович, редактировал(а) Гордон Ирина Григорьевна сен 02, 2020

Базовые показатели – основные показатели данных систем, предпочтительные для использования, т.к. для них выстраивается процесс управления качеством данных (data governance), в том числе ведется бизнес-гlossарий и организуется контроль качества.

Формируются следующим образом:

- Аддитивные и полуддитивные показатели рассчитываются в аналитических витринах на основе данных систем-источников, иногда при этом в качестве слагаемых используются другие базовые показатели.
- Неаддитивные показатели рассчитываются в OLAP-кубах на основе других базовых показателей. Расчет неаддитивных показателей в аналитических витринах нецелесообразен, т.к. в витринах данные хранятся на самом детальном уровне, а суммирование неаддитивных показателей некорректно.

Тезаурус
дата-терминов

Выручка

Создатель Викров Андрей Александрович, отредактировано только что

Бизнес-определение

Сумма денежных средств, причитающихся к получению компанией за продажу продуктов (товаров, услуг или

Аддитивный показатель в домене Продажи.

| | |
|-------------|---|
| Синонимы | |
| Не путать с | <ul style="list-style-type: none"> Сумма продажи |

Алгоритм расчета

Методология расчета привязана к атрибуту Группа направлений (товары, работы, услуги):

| Группа направлений | Правило расчета |
|--------------------|-----------------|
| Товары | |

Гlossарий
базовых объектов

Продажа

Создал(а) Викров Андрей Александрович, редактировал(а) Виктория Валерьевна мая 27, 2020

Продажа – сделка купли-продажи продукта с покупателем, оформленная выданным чеком.

Проданные продукты – продукты, по которым состоялась сделка купли-продажи.

См. также показатель сумма продажи

Тезаурус
бизнес-терминов

Из описания BI пользователь может переходить в статьи бизнес-гlossария за более подробной информацией

Выручка

Создатель Вихров Андрей Александрович, отредактировано только что

Бизнес-определение

Сумма денежных средств, причитающихся к получению компанией за **продажу продуктов** (товаров, услуг или платежей). Включает НДС. Измеряется в руб.

Аддитивный показатель в домене Продажи.

| | |
|-------------|-----------------|
| Синонимы | |
| Не путать с | • Сумма продажи |

Алгоритм расчета

Методология расчета привязана к атрибуту Группа направлений (товары, работы, услуги):

| Группа направлений | Правило расчета |
|--------------------|-----------------|
| Товары | ... |

Источники и стандартное представление

| Уровень | Объект | Название поля | Комментарий |
|---------------------------|---|---------------------------------|-------------|
| Стандартное представление | | | |
| OLAP-куб | olap_euroset Продажи | Выручка | |
| Источники | | | |
| Витрина | sandbox.dma.retail_transaction_tran_fn_vendor | financial_revenue_indicator_amt | |
| Стейджинг | ... | ... | |
| Источник | ... | ... | |

Прочие сведения

| | |
|-------------------------------|--|
| Владелец показателя | Владелец домена BI Продажи. Перечень доменов и их владельцев см. здесь . |
| Задачи по развитию показателя | <input checked="" type="checkbox"/> BIMS-14859 - Маржа и выручка подкатегории Кредит DEV DONE |

Описания базовых объектов – кирпичики бизнес-гlossария

В настоящий момент у нас порядка 70 законченных описаний и примерно столько же в разной степени готовности

Примеры - «средний чек», «количество активных ТТ», «выручка».

Тезаурусы – словарь языка данных

Тезаурус бизнес-терминов



Бизнес-термины, для которых не нашлось базового объекта (показателя, атрибута)

Сейчас в этом тезаурусе порядка 20 терминов, например «опердень», «покупатель», «продажа»:

Продажа

Создал(а) Вихров Андрей Александрович, редактировал(а) Виктория Валерьевна мая 27, 2020

Продажа – сделка купли-продажи продукта с покупателем, оформленная выданным чеком.

Проданные продукты – продукты, по которым состоялась сделка купли-продажи.

См. также показатель сумма продажи

Тезаурус дата-терминов



Термины, используемые при описании данных

Сейчас в этом тезаурусе порядка 50 терминов, например «витрина», «атрибут», «показатель»:

Показатель

Создал(а) Вихров Андрей Александрович, редактировал(а) Ирина Григорьевна авг 07, 2020

Показатель – характеристика операций или состояния бизнес-объекта в численном выражении. Например: выручка, запас на конец периода, цена.

Может рассчитываться как в аналитической витрине Тоторо (базовый показатель), так и в других дата-системах.

Синонимы

- Метрика
- Мера
- Факт
- Measure

Пример шаблона описания показателя /фрагмент/

Шаблон описания показателя

Бизнес-определение

В рамках базового определения: краткое, но полное. Это означает, что все необходимые понятия указываются, но не расшифровываются. Не нужно указывать название самого показателя. Указать единицу измерения и наличие НДС.

ОБРАЗЕЦ Сумма продаж, делённая на количество чеков. Измеряется в руб с НДС.

Указываем в явном виде тип объекта (показатель или атрибут), указываем домен со ссылкой на базовую страницу домена

ОБРАЗЕЦ Аддитивный показатель в домене Продажи.

Конкретизация понятий, которые использовались в определении. Если считаем, что понятие будет использоваться больше чем в одном показателе, создаём под это понятие отдельную статью.

Указываем подробности реализации конкретного объекта. В т.ч. описываем то, что не является "интуитивно-понятным" либо неоднозначным.

| | |
|-------------|--|
| Синонимы | |
| Не путать с | |

"Синонимы" - указываем разные названия одного и того же объекта, а также сленга, реально используемого в бизнесе

"Не путать с" - указываем названия других объектов, с которыми можно перепутать данный объект.

Алгоритм расчета

Указывается алгоритм расчета показателя (при наличии) со ссылками на показатели - элементы формулы

Период по умолчанию

Указываем временную метрику по умолчанию

По умолчанию период в отчётах определяется на основе даты выдачи чека (дата чека). В OLAP-кубе Продажи - это измерение Даты.

Технические подробности

Указываем детали из анализа кода, которые нужно указать, но которые не являются значимыми для бизнеса и затрагивают какие-то редкие случаи

› Раскрыть

Принципы бизнес-гlossария

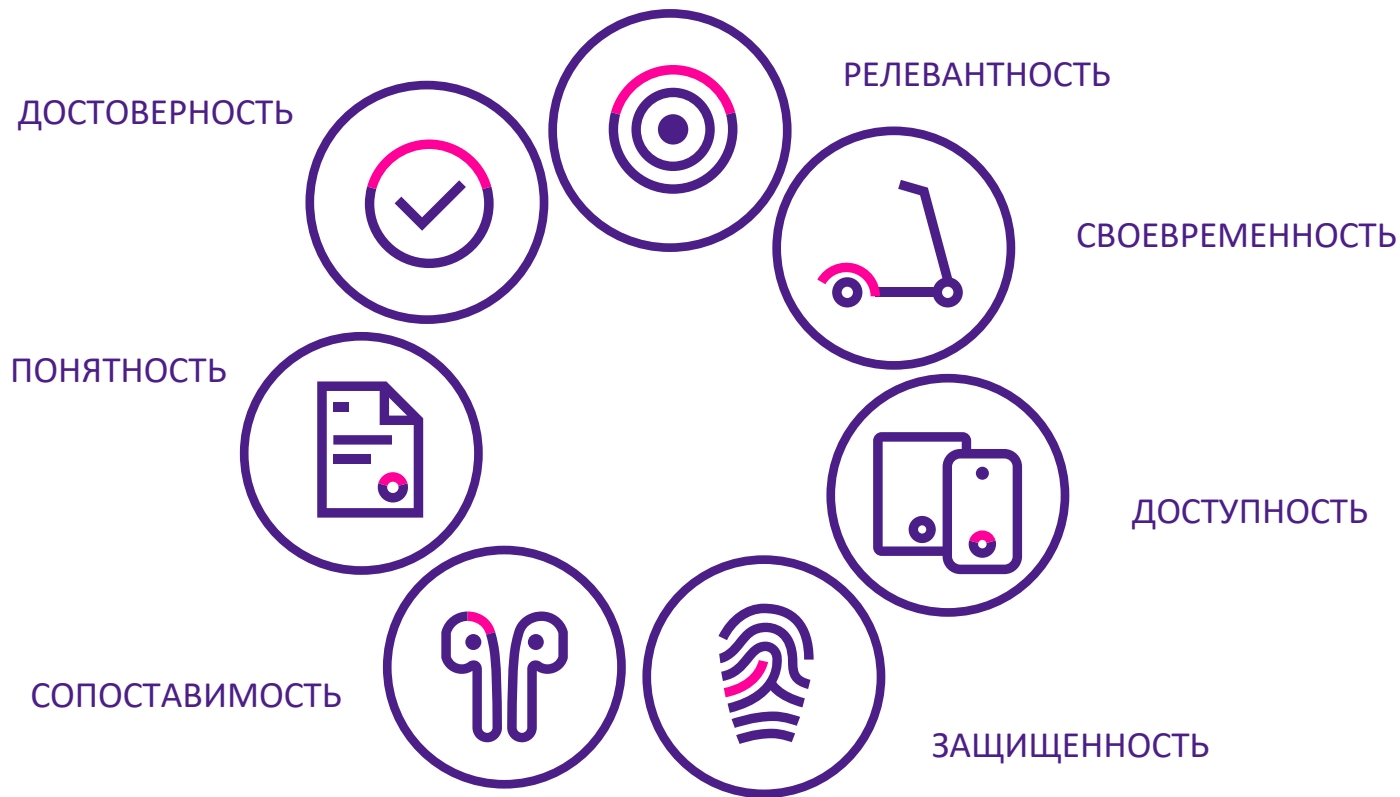
| ПРИНЦИП | ОПИСАНИЕ |
|---|--|
| СТАНДАРТИЗАЦИЯ И ШАБЛОНИЗАЦИЯ | Максимально используются шаблоны, стандарты содержания и оформления |
| ЕДИНАЯ ТЕРМИНОЛОГИЯ И ОДНОКРАТНОЕ ОПИСАНИЕ | Описания делаются единожды (в виде терминов глоссария). Повторно используются с помощью гиперссылок или зеркалирования текста макросами |
| БИЗНЕС-СМЫСЛ И ТЕХНИЧЕСКИЕ ОБЪЕКТЫ | Отображается бизнес-смысл данных, в привязке к техническим объектам, в которых они реализованы |
| ПРИОРИТЕТ БАЗОВЫХ ПОКАЗАТЕЛЕЙ | Основной акцент на описании базовых показателей и их дальнейшем использовании как готовых "кирпичиков" при описании всех остальных объектов |
| ПРОСТОТА И КРАТКОСТЬ | Описания формируются максимально просто и кратко, но без потери смысла |
| РЕЛЕВАНТНОСТЬ И СУЩЕСТВЕННОСТЬ | Наиболее важные сущности описываются в первую очередь и более глубоко, м енее важные - поверхностно |
| МЕСЕ ПРИНЦИП | Описания должны быть МЕСЕ ("взаимно исключающие, совместно исчерпывающие") |

Бизнес-выгоды бизнес-гlossария

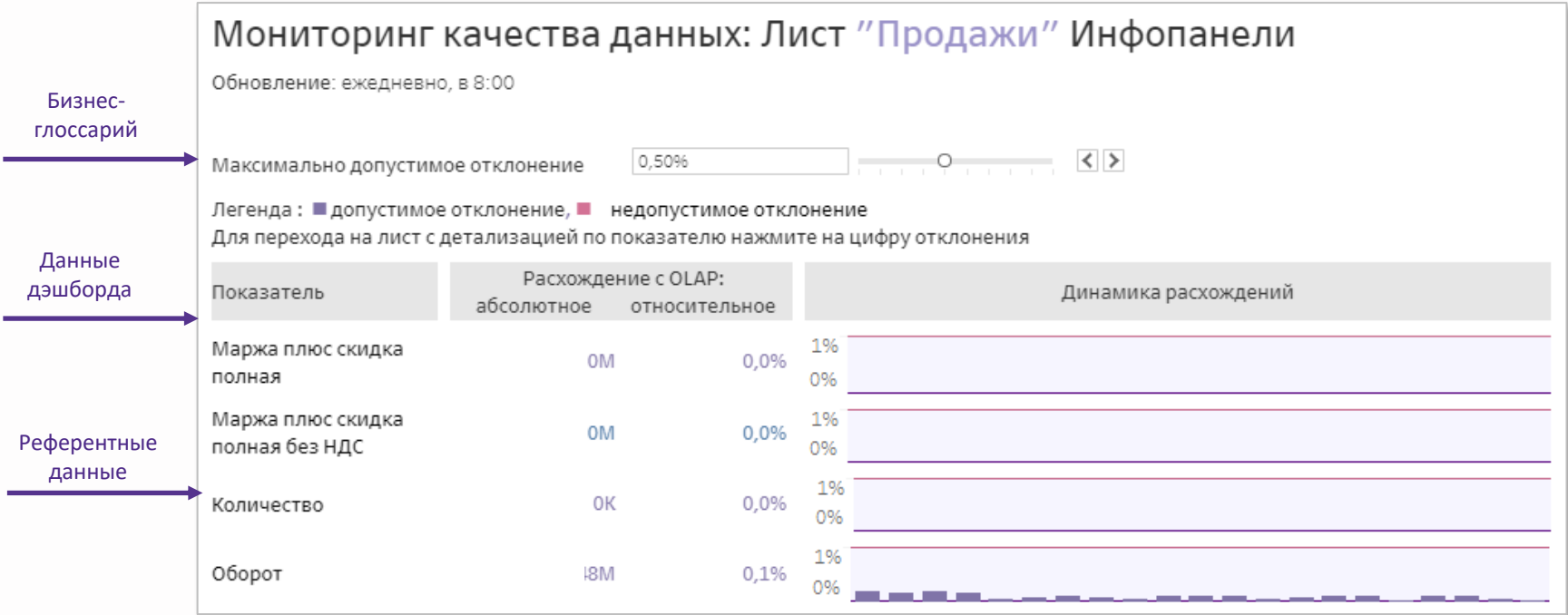


МОНИТОРИНГ КАЧЕСТВА ДАННЫХ

Критерии качества - и как их можно измерять



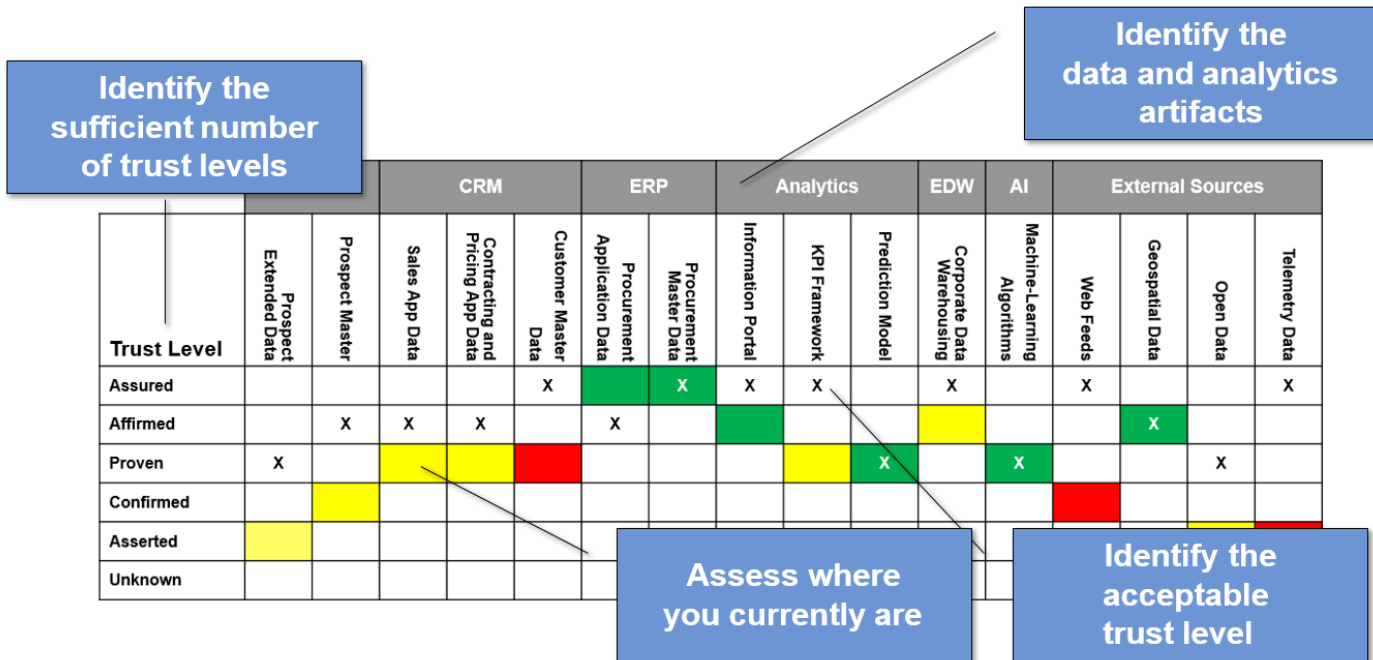
Дашборд по мониторингу качества данных



Модель оценки уровня доверия к данным

На основе анализа можно проводить разделение источников на золотые, серебряные, бронзовые, а также выбирать приоритеты

Common Trust Model



ИНСТРУМЕНТАРИЙ

Наши текущий инструментарий

Используем Confluence в качестве платформы для бизнес-гlossария и репозитория основных метаданных

Плюсы +++

1. доступность
2. гибкость
3. кастомизируемость
4. гиперлинки
5. отслеживание версий
6. возможность поиска
7. интеграция с Jira

Минусы ---

1. нет интеграции с DWH
2. нет lineage
3. нет преднастроенных шаблонов
4. нет потока операций

Workarounds

3. делаем стандарты, шаблоны
4. используем Jira, папки в Confluence

Специальные инструменты для бизнес-гlossария

Просмотр информации о термине

IBM Infosphere

Термин

Значение остатка по счету

Определение

Категория

Пользовательские атрибуты

| | |
|---------------------------|------------------------------------|
| Полное описание | Значение остатка по счету |
| Родительская категория | Бизнес-классификатор > Бухгалтерия |
| Состояние | Кандидат |
| Общая информация | |
| Номер атрибута | 40040015 |
| Тип атрибута | Финансовый 23 |
| Целевая витрина. Ресурс | + |
| Целевая витрина. Кредиты | + |
| Модификатор | False |
| Тип | Нет |
| Created By | isadim admin |
| Дата создания | 06.04.2015 14:44:04 |
| Last Modified By | isadim admin |
| Дата последнего изменения | 17.04.2015 15:22:35 |

Связанные термины

Назначенные активы (14)

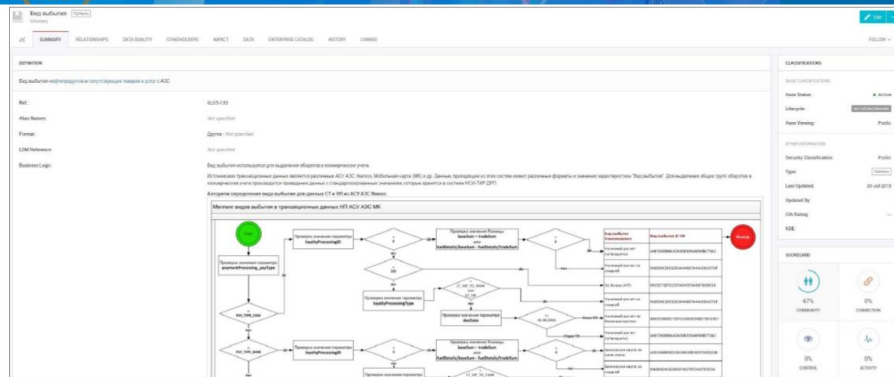
| |
|------------------------------|
| 11-14x14 |
| Posted Amount / Atomic W |
| Posted Amount Domestic / |
| Исходящий остаток в рублях / |
| Факты |
| Исходящий остаток в рублях / |
| Факты |

Связанные технические метаданные



- ведение паспортов бизнес-сущностей, определений, методик расчета
- привязка к владельцам, стюардам и т.п.
- привязка к бизнес-процессам, тегам
- ведение истории изменений

Бизнес-гlossарий: основной функционал (Axon)



| Start | Update Time | Field | From | To | Author | Date |
|----------|---------------|-------------|--|--|-----------|------------|
| Category | Updated | Description | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Сотрудник | 11.04.2015 |
| Category | Status Change | Category | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Сотрудник | 11.04.2015 |
| Category | Updated | Description | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Сотрудник | 11.04.2015 |
| Category | Updated | Field | 30 | 30 | Сотрудник | 11.04.2015 |
| Category | Added | Role | | | Сотрудник | 11.04.2015 |
| Category | Status Change | Category | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Нарушил Типы из группы "Нарушения системы ИТ-операций" | Сотрудник | 11.04.2015 |
| Category | Added | Role | Admin Status | | Сотрудник | 11.04.2015 |

Informatica Axon

New Business Term

MS Azure

Term Name

enter a term name

Parent Term

No parent term available

Definition

add the business definition for the term

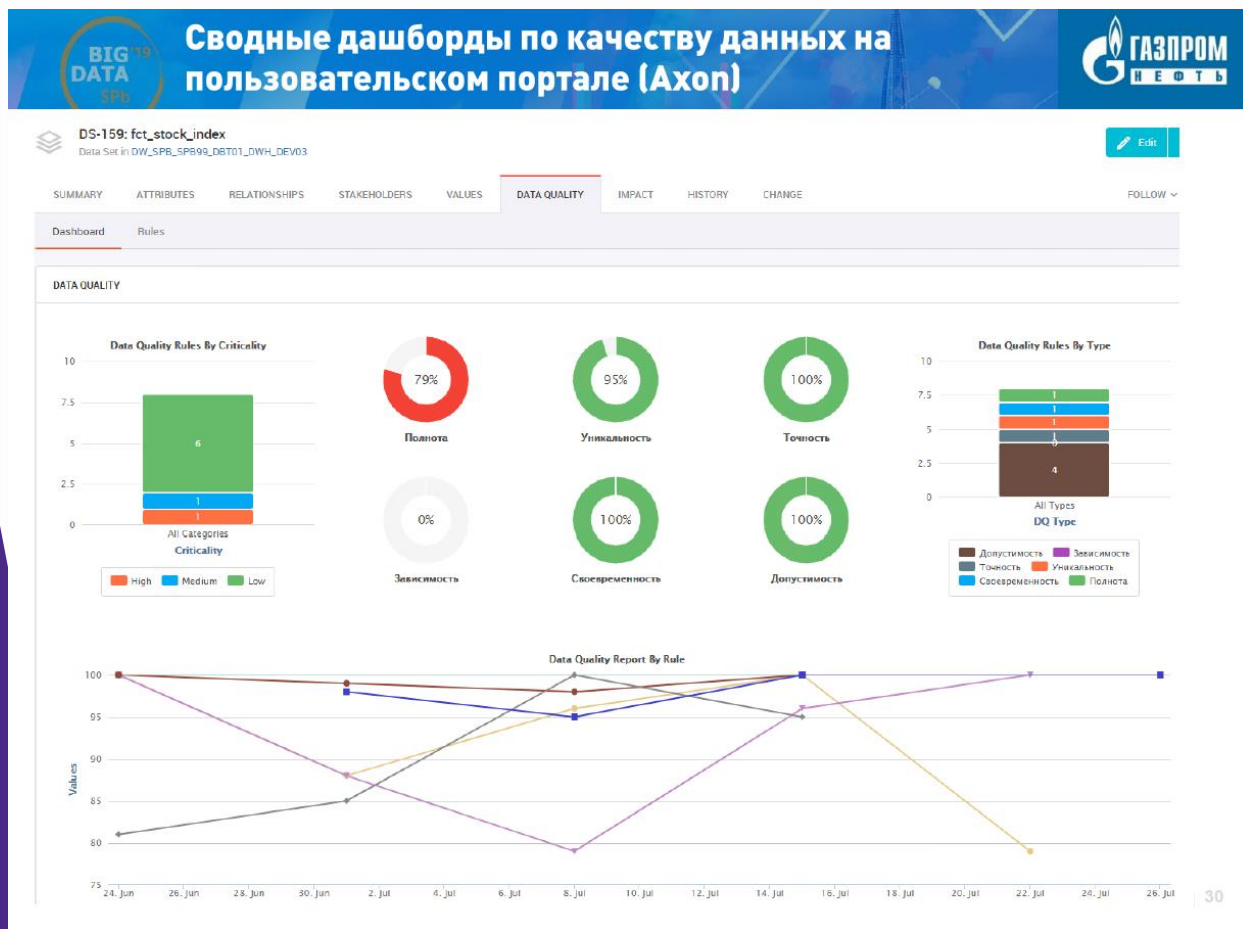
ess rule, etc.

te and New

Create

Cancel

Реализация дэшбордов по качеству в специальном инструменте /Informatica Axon/



Компоненты продукта класса information stewardship

Critical Components



- Packaged as One Solution
- The User Interface/User Experience Is for Business Users, Not IT Users

Рекомендации по выбору инструмента

Recommendations

- ✓ Define the role of information steward for your information governance program before you select information stewardship applications or solutions.
- ✓ Evaluate the capabilities of your existing data management tools, including data quality, metadata management and federation/integration capabilities, before buying an information stewardship application or solution.
- ✓ Run a proof of concept of vendor solutions involving all contributing roles, such as business users, information governance board members, information architects, information stewards and business analysts.
- ✓ For a new solution, balance your solution functionality requirements with other aspects such as vendor execution and vision, service and support, and total cost of ownership.
- ✓ When addressing the information stewardship use case, pay close attention to all the relevant dimensions (people, process and technology) that are relevant for an effective use of the solution to maximize return on investment through reuse, while minimizing administrative costs and errors due to inconsistencies across technologies.



Спасибо!



vikhrov.andrey@svyaznoy.ru



[@anvih](https://t.me/anvih)