

# Системы управления базами знаний в контексте управления данными (взгляд из практики)

Докладчик: к.т.н., доц. Алексей Незнанов

2025-01-16

# О докладчике

- Алексей Андреевич Незнанов
  - К.т.н., доцент
  - Член IEEE, РАИИ, НБМЗ
  - *Senior Data Scientist* в компании Шлюмберже
  - Старший научный сотрудник международной лаборатории интеллектуальных систем и структурного анализа ФКН НИУ ВШЭ (<http://cs.hse.ru/ai/issa>)
  - Консультант «маленького гида по большим данным» на Постнауке ([http://postnauka.ru/author/a\\_neznanov](http://postnauka.ru/author/a_neznanov))
  - Автор учебника, учебных пособий, 12 авторских курсов и более 65 научных публикаций



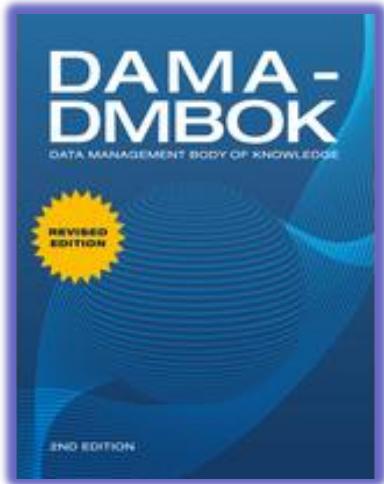
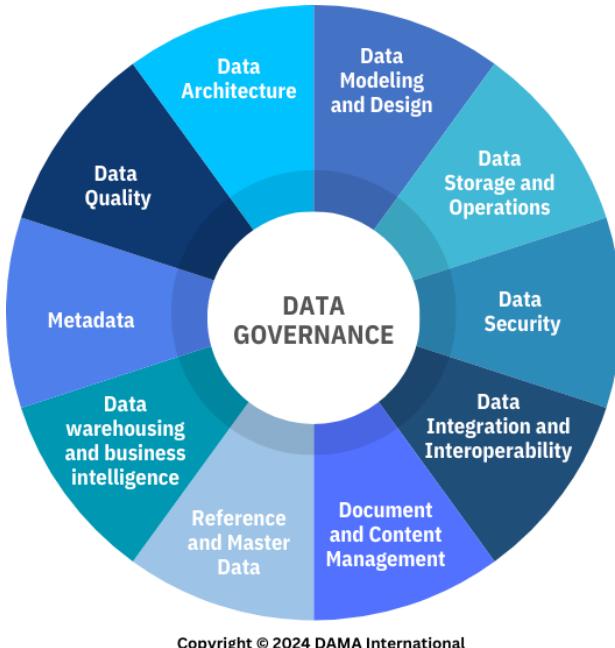
# Один из постулатов обработки данных

## «Мусор на входе – мусор на выходе» ("Garbage in – garbage out")

- Качество данных определяет предельное качество их обработки
  - Особенno при использовании комплексных процедур анализа данных и машинного обучения
- Управление данными
  - **DataOps!** Вместе с *DevOps* и более узкими *MLOps*, широкими *AIOps* и т.п.
    - Акцент: управление мастер-данными (**MDM**) и управление качеством данных (**DQM**)
  - «Интеллектуализация» обработки данных
- Хоть чуть-чуть работы с «явным знанием» (sic!)
  - Базовая онтология предметной области и интерпретация результатов анализа данных
    - Связь с мастер-данными и метаданными хранилищ данных
  - Подходы к гармонизации данных:
    - Идентификация, дедупликация, онтологизация, актуализация, версионирование, ...
  - **Гибридизация систем ИИ**

# Общая рамка – управление данными (*DataOps*)

- *DAMA-DMBOK2: Data Management Body of Knowledge. Revised Edition, Technics Publications, 2024. 626 p.*
  - Свод знаний (ВоК)! Готовится третье издание
    - Есть неплохой русский перевод (не Revised!)
  - *DAMA International* (<http://dama.org/content/body-knowledge>)





## Данные и их обработка

- ✓ Что такое данные?
- ✓ Откуда берутся «наборы данных» и каковы основные свойства «источников данных»?
- ✓ Почему при обсуждении данных невозможно обойтись без многоуровневых метаданных, а также интерфейсов и протоколов?
- ✓ Зачем хранилищам данных нужны витрины?
- ✓ Можно ли оценить качество данных? И какие критерии для этого используются?

# Информация и данные

- **Информация [information]**
  - Это базовое понятие, не определяемое конструктивно без использования рекурсии из-за проблемы герменевтического круга
    - От лат. *informatio* – разъяснение, изложение, осведомленность
  - Сведения, воспринимаемые человеком и (или) спец. устройствами как отражение фактов материального или духовного мира в процессе коммуникации (ГОСТ 7.0-99)
  - И обилие других неконструктивных определений
    - Моё любимое: **информация – это не материя и не энергия!** (Н. Винер)
- **Данные [data]** – любая информация, представленная в форме, пригодной для хранения, передачи и обработки средствами вычислительной техники
- Информационные процессы, использующие данные, называются **обработкой данных [data processing]** (и изучаются информатикой 😊 – гримасы терминологии)

# Задачи обработки данных

- Данные для нас являются тем, что обрабатывают **информационные системы (ИС)**
- Обработка данных разбивается на несколько областей
  1. **Ввод/вывод** – извлечение информации из внешней среды и её формализация в виде данных или сохранение данных во внешней среде
  2. **Преобразование** форматов – изменение способа представления данных
  3. **Телекоммуникации** – передача данных между компонентами распределённых информационных систем
  4. **Хранение** – обеспечение долговременной доступности, целостности и защищенности данных
  5. **Поиск** данных – выполнение поисковых запросов к массивам данных
  6. **Переработка** – существенное преобразование содержания и/или формы данных, выполняемое на основе анализа и/или синтеза

# Стандартная классификация видов данных с точки зрения бизнеса

## Данные (+их метаданные)

«Чужие»

Справочные  
(референсные)  
данные

Гармонизируемые  
данные

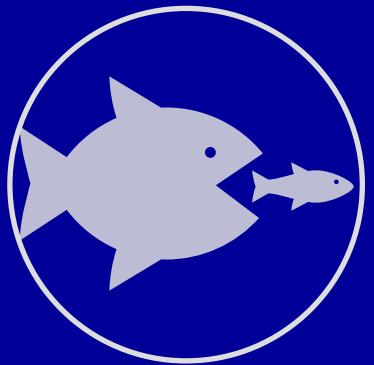
«Свои»

Мастер-данные

Оперативные  
данные

Исторические  
данные

# Основные проблемы и аспекты управления данными



## Хранение и передача

- Ресурсы хранения
- Каналы связи
- Распределение



## Эффективная обработка

- Вычислительные ресурсы
- Распараллеливание



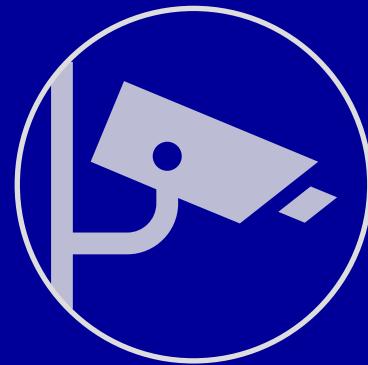
## Возможности гармонизации

- Идентификация сущностей
- Общие справочники
- Ссылки
- Контекст



## Актуальность и версионирование

- Отметки времени в едином формате
- Корректность отметок времени на одной временной оси



## Особенности владения и использования

- Информационная безопасность
- Персональные данные! См. ФЗ152



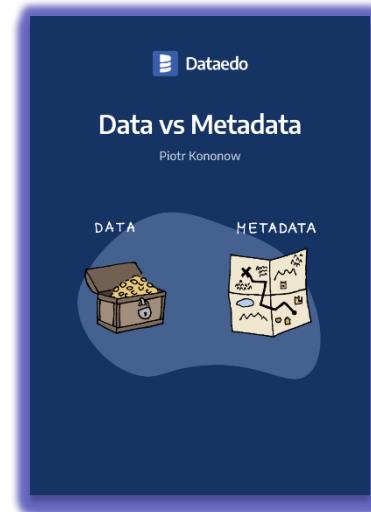
## Режимы доступа

- Аудит
- Комплаенс
- Цитирование данных
- Открытые данные [Open Data]



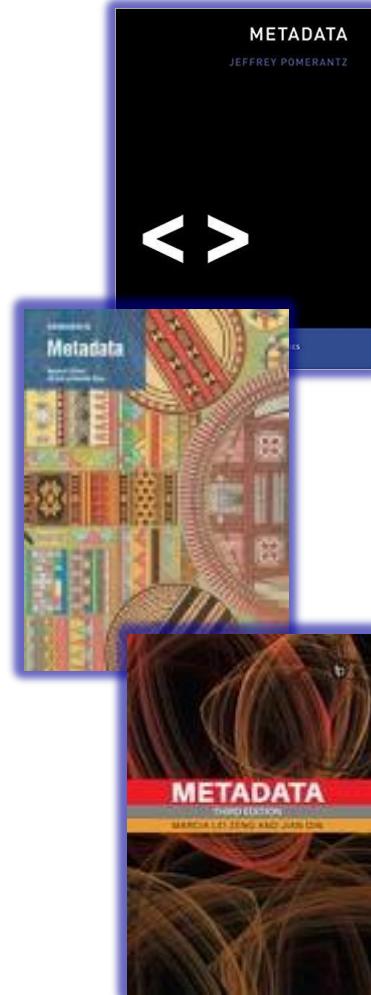
# Данные и метаданные

- Знания появляются на уровне **метаданных** [*metadata*]
  - Данные, описывающие другие данные 😊
    - UNC – *Metadata for Data Management: A Tutorial. Standards/Schema* (<http://guides.lib.unc.edu/metadata/standards>)
    - Dataedo – *Data vs Metadata* (<https://landing.dataedo.com/data-vs-metadata>)
- Управление метаданными
- Формализация метаданных
  - Без формально описанных метаданных работать с данными не получится!
- Метаданные для метаданных (бесконечная иерархия)
  - Где остановиться? **Метаданные**, **метаметаданные**, **метаметаметаданные**, ...
  - Проблема согласования онтологий и онтиков
  - Проблема интероперабельности – решается!
- Моделирование
  - UML (<http://www.uml.org>), BPMN (<http://www.bpmn.org>), IDEF, ...
  - Archimate? ...



# Введение в область

- *Pomerantz J. Metadata. MIT, 2015*
  - <http://www.amazon.com/Metadata-MIT-Press-Essential-Knowledge/dp/0262528517>
- *Introduction to Metadata. Edited by Baca M. 3<sup>rd</sup> ed. 2016*
  - <https://www.getty.edu/publications/intrometadata/>
- *Zeng M.L., Qin J. Metadata. 3<sup>rd</sup> ed. Facet Publishing, 2022*
  - <https://www.routledge.com/Metadata/Zeng-Qin/p/book/9781783305889>



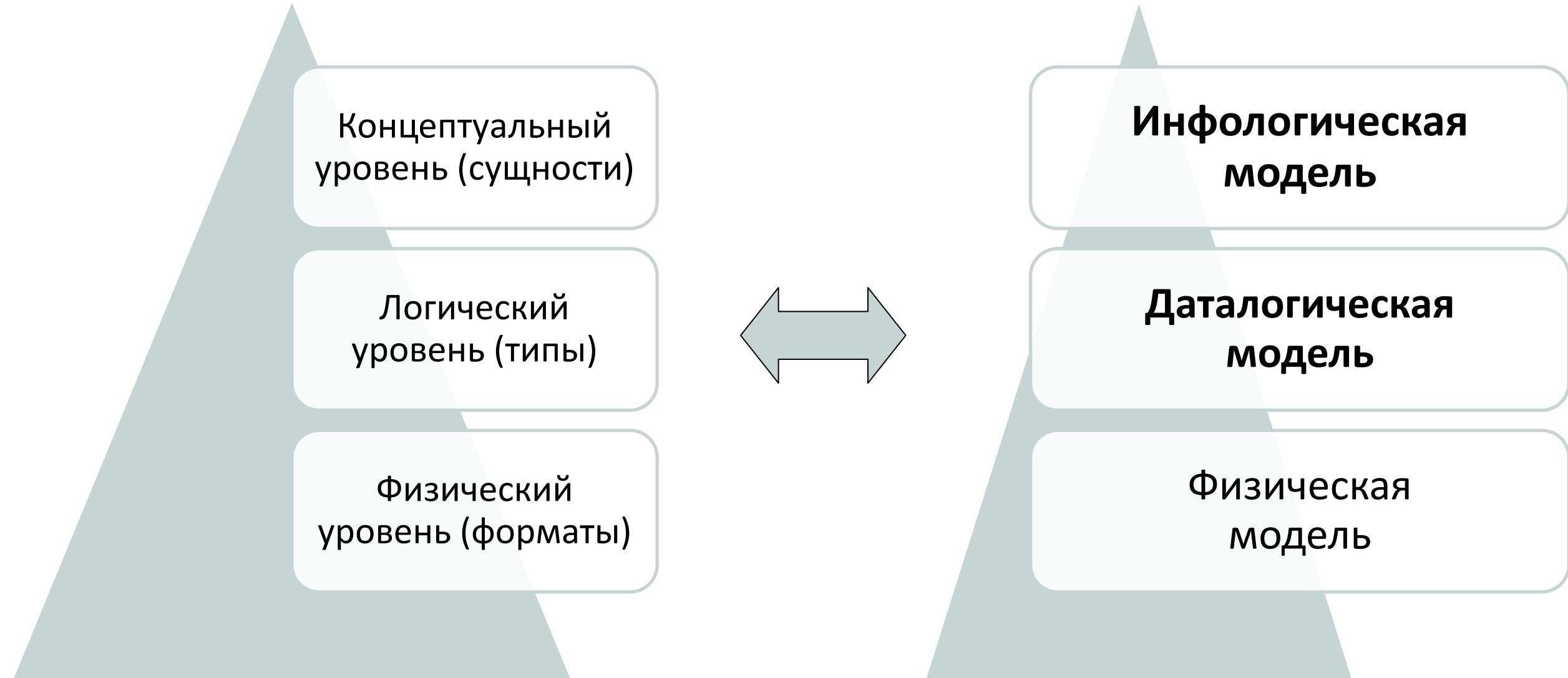
# Информационные артефакты и документы

- **Артефакт** (от лат. *arte* – искусственный и *factus* – сделанный) – искусственно созданный с некоторой целью материальный или нематериальный (информационный) объект
  - Нас собственно интересуют информационные артефакты с их **носителями**
- **Документ** [*document*] – артефакт, являющийся средством закрепления различным способом на материальном носителе сведений о фактах, событиях, явлениях объективной действительности и мыслительной деятельности человека
  - Документы принято условно разделять на:
    - **Первичные** (содержащие запись *фактов*)
    - **Вторичные** (содержащие результаты *анализа и интерпретации* этих фактов)
  - У документов есть нормативные метаданные – **реквизиты!**

# Формат данных как базовые метаданные

- **Формат данных [data format]** – совокупность правил представления и интерпретации данных в памяти компьютера, на внешних носителях, при операциях ввода/вывода и при передаче по каналам связи
  - Главное: формат определяет способ **кодирования** данных в некоторой **знаковой системе**
    - То есть формат – это описание логической/физической структуры информационного элемента
- **Основные форматы атомарных значений:**
  - Числа: **IEEE 754-2008 “IEEE Standard for Floating-Point Arithmetic”** (<http://grouper.ieee.org/groups/754>)
  - Строки: **UNICODE** (<http://www.unicode.org>)
  - Дата/время: **ISO 8601:2019 “Data elements and interchange formats – Information interchange – Representation of dates and times”** (<http://www.iso.org/iso/ru/home/standards/iso8601.htm>)
  - ...

# Трёхуровневая модель абстракции данных

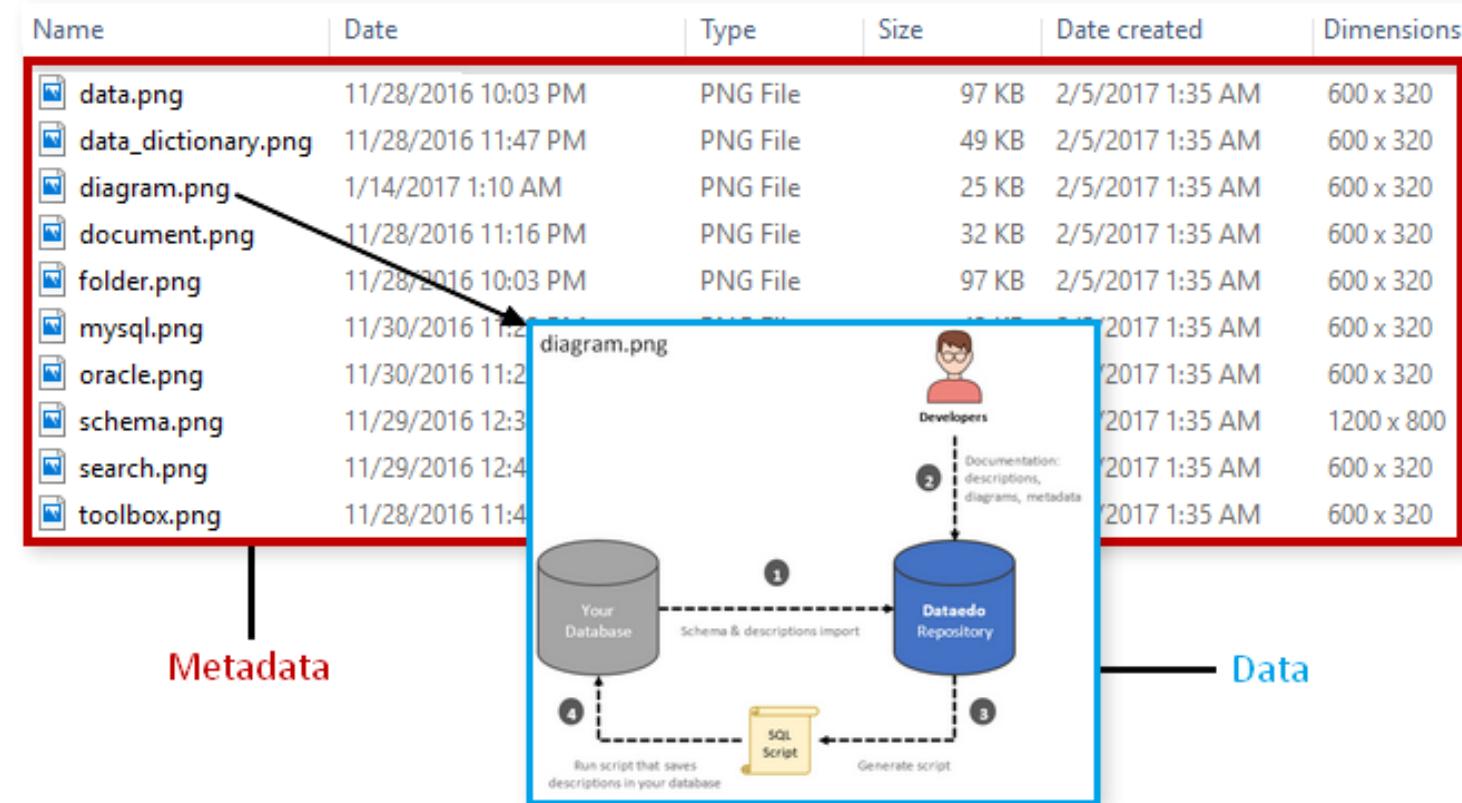


# Описание документов как следующий уровень

- Следующий уровень – реквизиты содержимого
- Пример Dublin Core:
  - ISO 15836-1:2017 *Information and documentation – The Dublin Core metadata element set – Part 1: Core elements* (<http://www.iso.org/obp/ui/#iso:std:iso:15836:-1:ed-1:v1:en>)
  - ISO 15836-2:2019 *Information and documentation – The Dublin Core metadata element set – Part 2: DCMI Properties and classes* (<http://www.iso.org/obp/ui/#iso:std:iso:15836:-2:ed-1:v1:en>)
    - Naden C. International Standard for Descriptive Metadata Just Update. 2020 (<http://www.iso.org/news/ref2474.html>)
- Пример ГОСТ Р 7.0.97-2016 (заменил ГОСТ Р 6.30-2003)
  - Стандарт вводит 30 базовых реквизитов  
(от «подписи» и «грифа» до «Государственного герба Российской Федерации»)

# Форматы файлов (1) – файловые системы

- Файл как элемент файловой системы
  - Атрибуты файлов
  - Поиск по атрибутам



# Форматы файлов (2) – содержимое

- Файлы, с которыми могут работать конкретные программы

- *Sustainability of Digital Formats: Planning for Library of Congress Collections – Format Descriptions* (<https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>)

**Sustainability of Digital Formats: Planning for Library of Congress Collections**

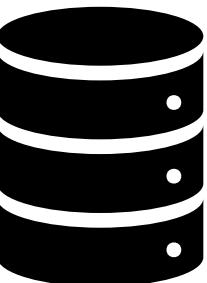
[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

[Format Descriptions](#) >> Format Description Categories >> [Browse Alphabetical List](#) >> [Format Descriptions as XML](#)

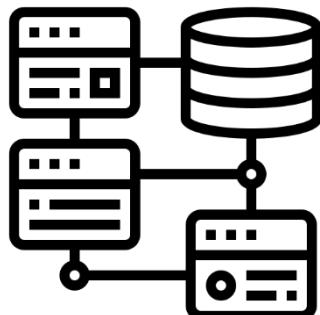
**Format Descriptions**

<b>Still Image</b> <ul style="list-style-type: none"><li>• <a href="#">SVG_1_1</a></li><li>• <a href="#">TIFF_6</a></li><li>• <a href="#">All still image format descriptions</a></li></ul>	<b>Sound</b> <ul style="list-style-type: none"><li>• <a href="#">WAVE</a></li><li>• <a href="#">MP3_FF</a></li><li>• <a href="#">All sound format descriptions</a></li></ul>	<b>Moving Image</b> <ul style="list-style-type: none"><li>• <a href="#">MPEG-4_FF_2</a></li><li>• <a href="#">AVI</a></li><li>• <a href="#">All moving image format descriptions</a></li></ul>
<b>Textual</b> <ul style="list-style-type: none"><li>• <a href="#">PDF/A_family</a></li><li>• <a href="#">DOCX/OOXML_2012</a></li><li>• <a href="#">All text format descriptions</a></li></ul>	<b>Web Archive</b> <ul style="list-style-type: none"><li>• <a href="#">ARC_IA</a></li><li>• <a href="#">WARC</a></li><li>• <a href="#">All Web archive format descriptions</a></li></ul>	<b>Datasets</b> <ul style="list-style-type: none"><li>• <a href="#">DBF</a></li><li>• <a href="#">HDF5</a></li><li>• <a href="#">All dataset format descriptions</a></li></ul>
<b>Geospatial</b> <ul style="list-style-type: none"><li>• <a href="#">ESRI_shape</a></li><li>• <a href="#">GeoPackage_1_0</a></li><li>• <a href="#">All geospatial format descriptions</a></li></ul>	<b>Generic</b> <ul style="list-style-type: none"><li>• <a href="#">ASF</a></li><li>• <a href="#">RIFF</a></li><li>• <a href="#">All generic format descriptions</a></li></ul>	

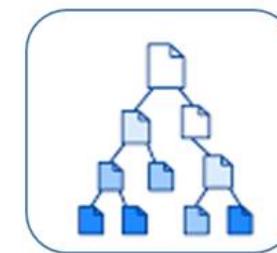
# Базы данных как торжество метаданных



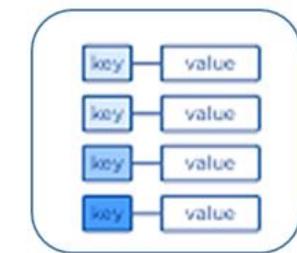
- **База данных (БД) [database (DB)]** – структурированная совокупность данных, организованная по единым правилам, включающим принципы описания, хранения и манипулирования этими данными
  - Как правило, БД является информационной моделью некоторой предметной области
  - Интеграция данных и метаданных (вплоть до некоторого уровня)
- **Системы управления базами данных (СУБД)** как основа всего дальнейшего



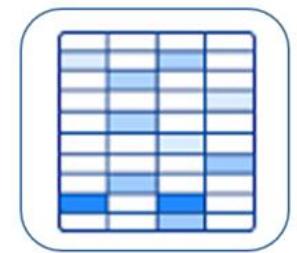
**SQL → NoSQL**



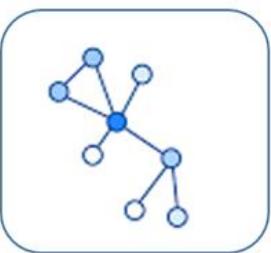
Document  
Store



Key-Value  
Store

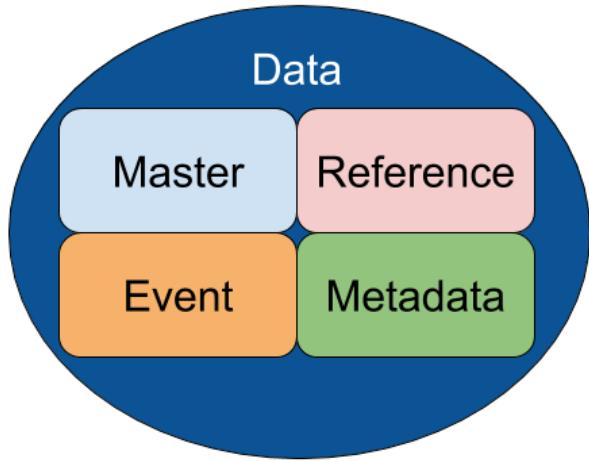


Wide-Column  
Store



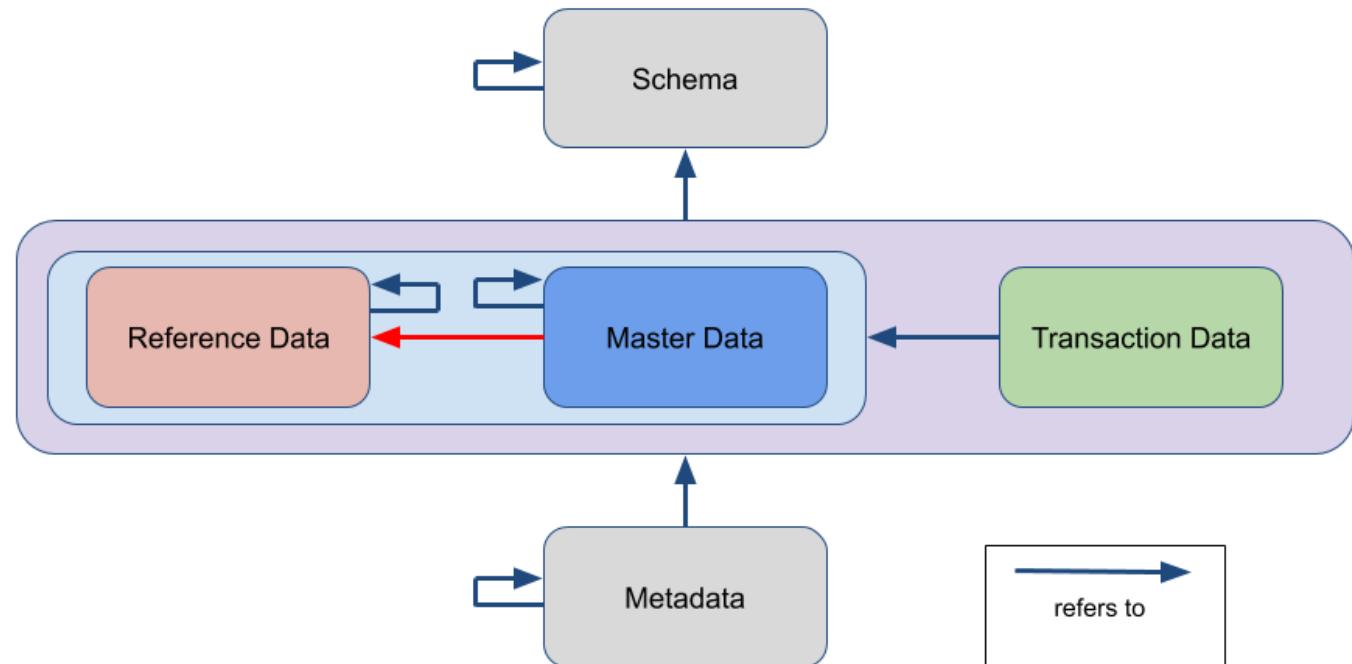
Graph  
Store

# Вариант Алана Фридмана



Depicting master, reference and other categories of data as peers

- Master
  - Customer
  - Product
- Reference
  - Country
  - Unit of Measure
- Event
  - User Action
  - Agronomic Activity
- Metadata
  - Resource Metadata
  - Process Metadata

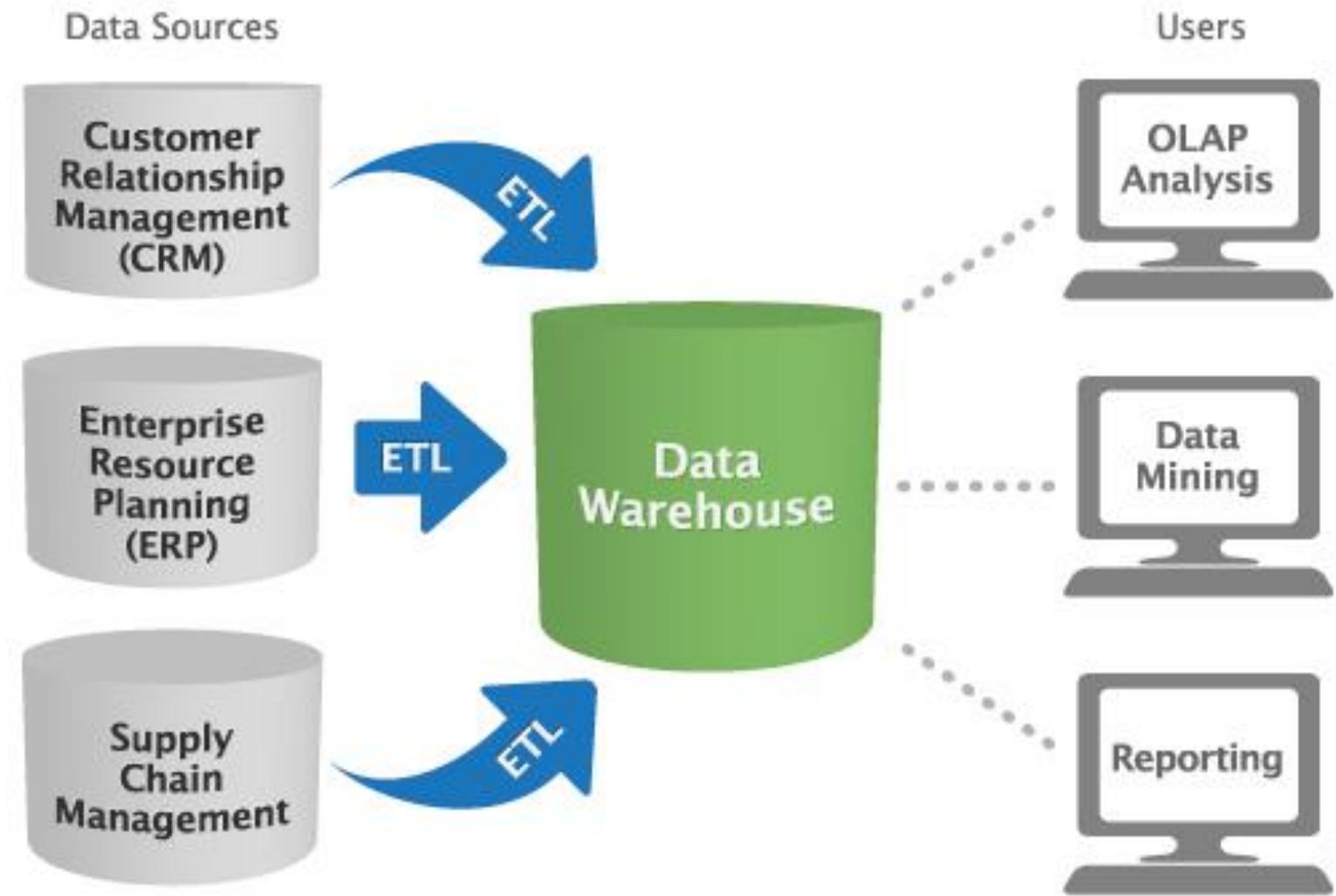


Master data has a one-way dependence on its master, reference data

# Разные взгляды на данные

- Классическая идеология «хранилищ данных» и «витрин данных»
  - Два главных (различных!) описания
    - *Inmon B. Building the Data Warehouse*
    - *Kimball R. The Data Warehouse Toolkit*
  - Современная терминология – с 1998 года! То есть уже более четверти века...
    - *Inmon B. Data Mart Does Not Equal Data Warehouse. DM Review, 1998*
- **Витрина данных [data mart]** –рез хранилища данных, представляющий собой структурированный массив тематической информации, ориентированный на решение конкретной задачи конкретным пользователем
  - Используется для управления доступом, упрощения визуализации данных и оптимизации работы с изменямыми со временем и темпоральными данными

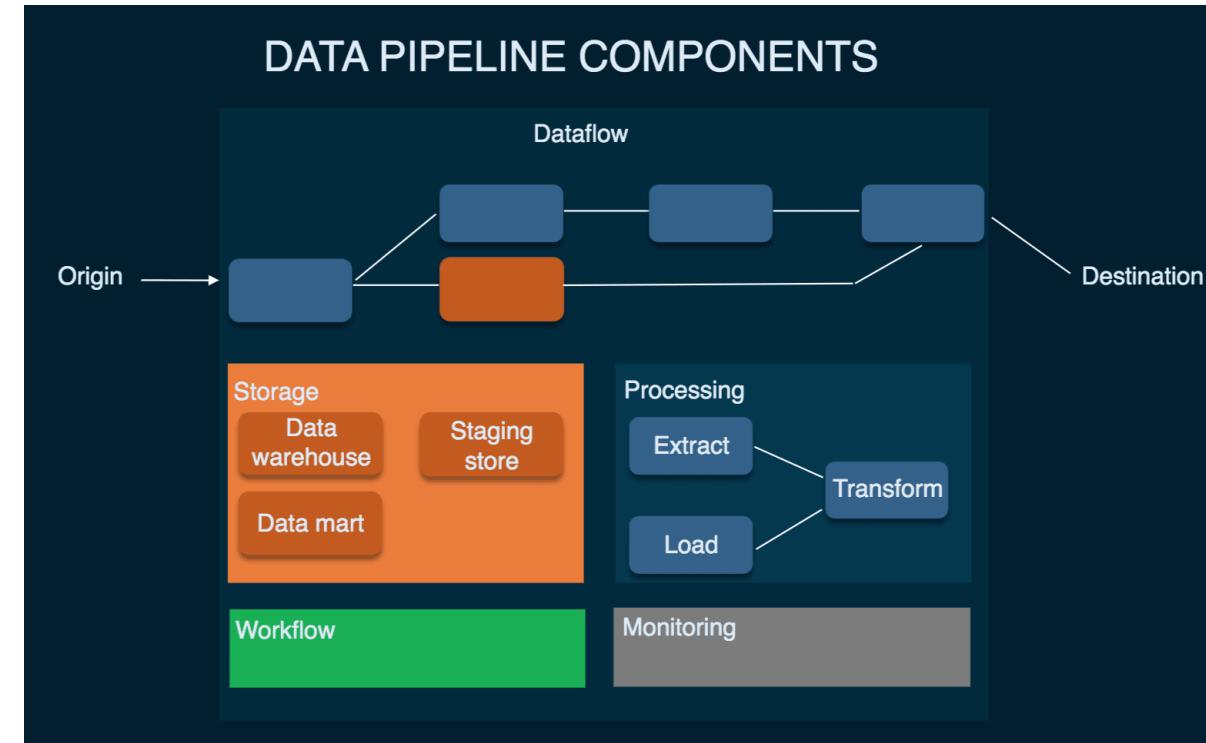
# Централизация на основе хранилища данных



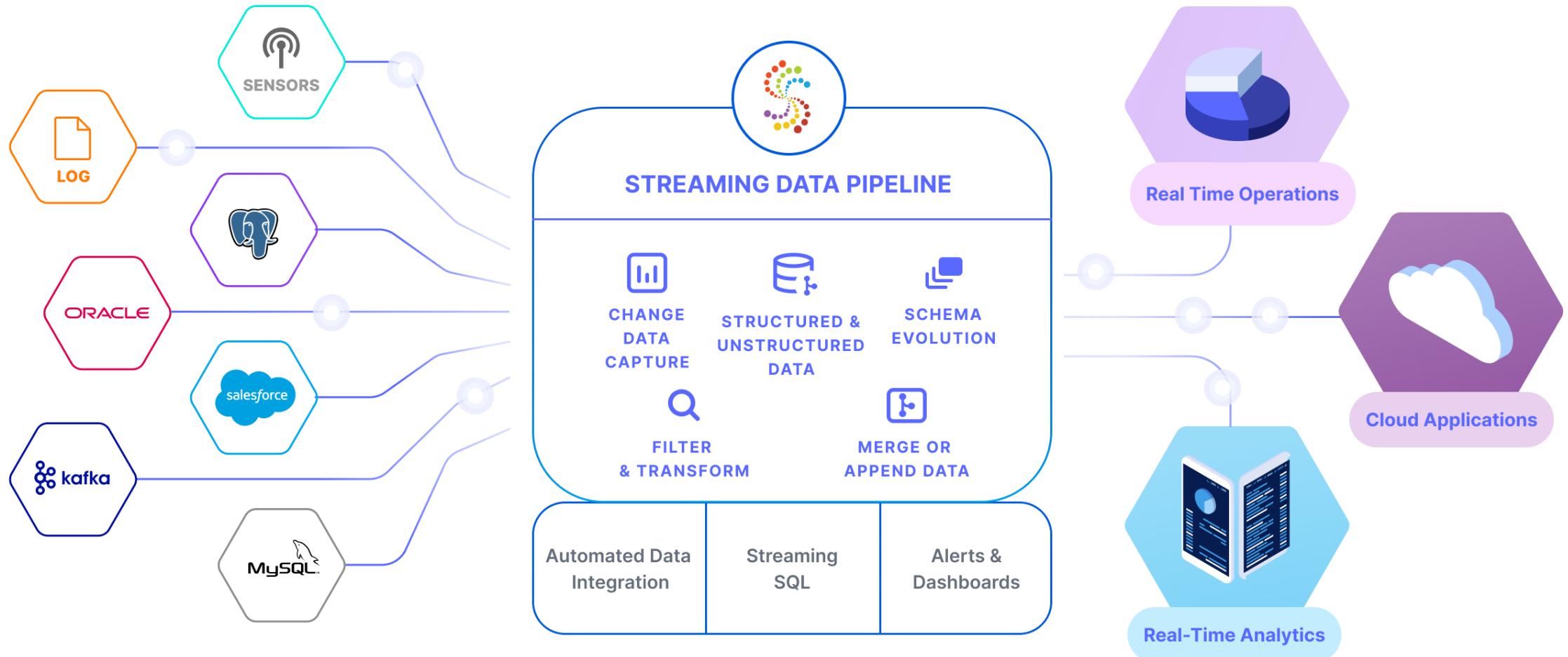
© Houston Neal, 2011

# Потоки данных и «датапроводы»

- Поток данных [*dataflow*] должен где-то «течь»
- «Датапровод» [*data pipeline*]!
  - Hevo Data – What is a Data Pipeline? Types, Components and Use Cases. 2022 (<http://hevodata.com/learn/data-pipeline/>)



# Как мы строим датапроводы?



John Kutay. What Is DataOps and How Can It Add Value to Your Organization?  
(<https://www.strim.com/blog/what-is-dataops-and-how-can-it-add-value-to-your-organization/>)

# *ETL – Extract/Transform/Load*

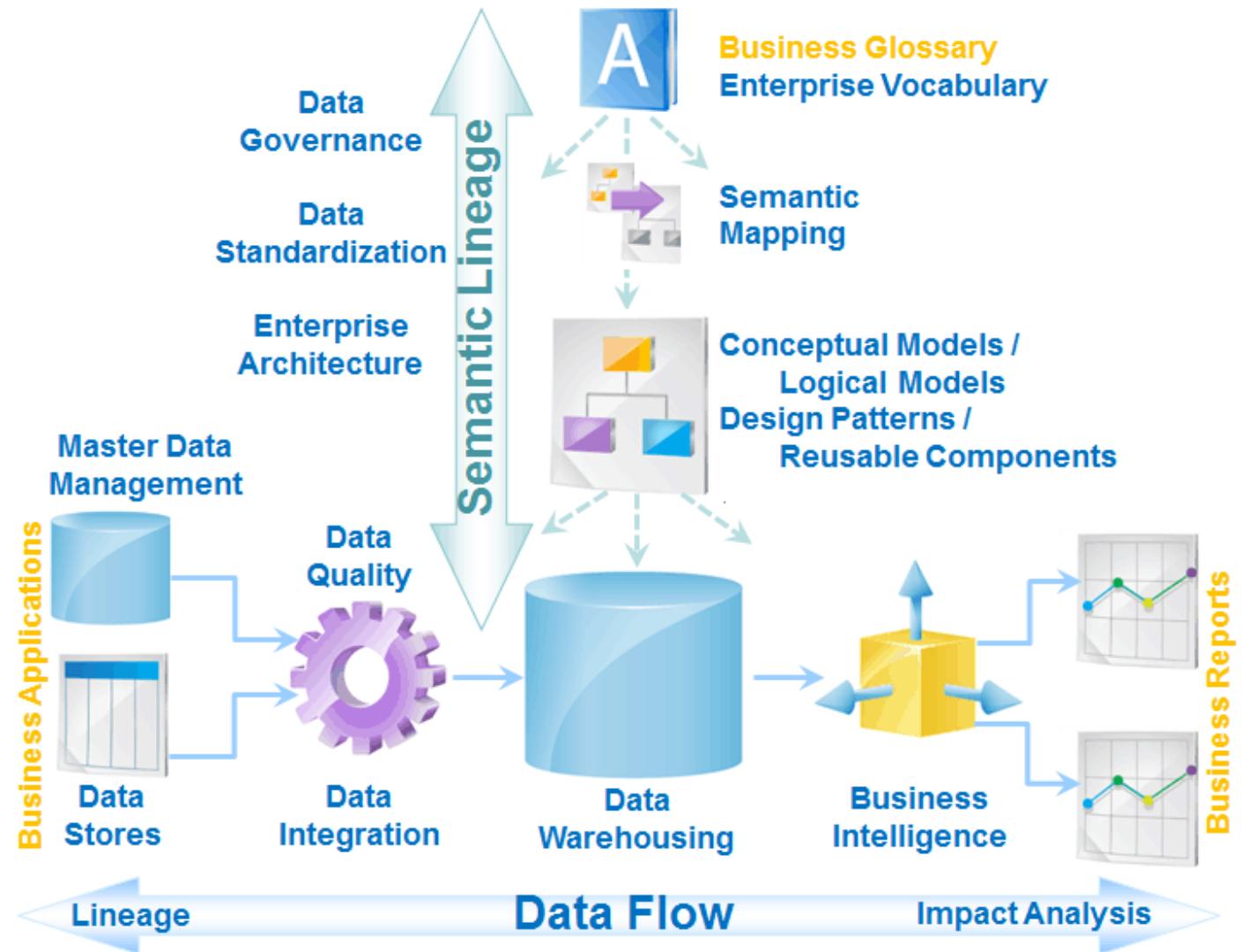
- Предобработка (препроцессинг) данных:
  - **Extract** – извлечение из внешних источников данных
  - **Transform** – трансформация, гармонизация и очистка для удовлетворения требованиям целевой системы
  - **Load** – загрузка в хранилище данных целевой системы
- Концепции:
  - *IBM Cloud Learn Hub – What is ETL (Extract, Transform, Load)?* (<http://www.ibm.com/cloud/learn/etl>)
  - *Microsoft Azure Data Architecture Guide – Extract, transform, and load (ETL)* (<http://docs.microsoft.com/en-us/azure/architecture/data-guide relational-data/etl>)
- *ETL vs ELT*
  - Где делаем трансформацию? На что при этом опираемся? Как формализуем?

# Иллюстрация управления всем сразу

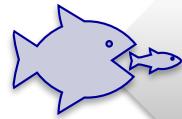
## ● Meta Integration

(<http://www.metaintegration.net>)

- Metadata Integration
- Metadata Harvesting
- Metadata Management
- Enterprise Architecture
- Data Governance



# Большие данные – пять букв «V» (5V)



Объём [*volume*]



Скорость обработки [*velocity*]



Многообразие [*variety*]

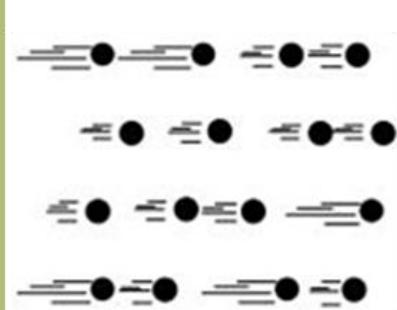
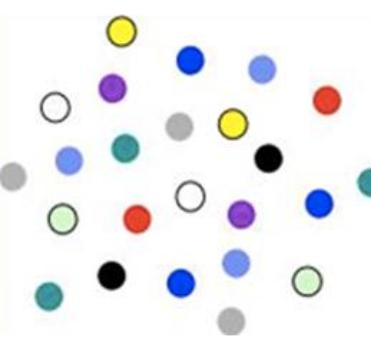
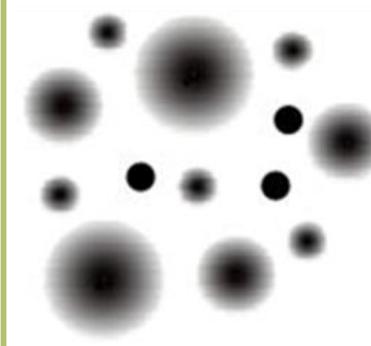


«Правдивость» выводов [*veracity*]



Извлечение «ценности» [*value*]

# Иллюстрация 5V

Volume	Velocity	Variety	Veracity	Value
				
<b>Data at Rest</b> Terabytes to Exabytes of existing data to process	<b>Data in Motion</b> Streaming data, requiring milliseconds to seconds to respond	<b>Data in Many Forms</b> Structured, unstructured, text, multimedia,...	<b>Data in Doubt</b> Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations	<b>Data into Money</b> Business models can be associated to the data

Adapted by a post of Michael Walker on 28 November 2012

# Интероперабельность – Понятие

- **Интероперабельность [interoperability]** – многозначное понятие, но нам интересно следующее определение:
  - Характеристика набора систем, интерфейсы которых известны (документированы), взаимозаменямо функционировать в качестве элемента надсистемы
    - Она шире, чем **совместимость** [*compatibility*]
    - Она также позволяет хорошо различать **уровни абстракции** и дополнить модель **открытых систем** (*OSI RM*)
    - Она позволяет точнее говорить о **платформах** [*platform*], как о том месте, где максимально проявляется
    - Именно с позиций интероперабельности удобно обсуждать интерфейсы, протоколы и **форматы данных**
- Чем выше уровень интероперабельности, тем быстрее развиваются сервисы платформы

Прикладной уровень

Уровень представления

Сеансовый уровень

Транспортный уровень

Сетевой уровень

Канальный уровень

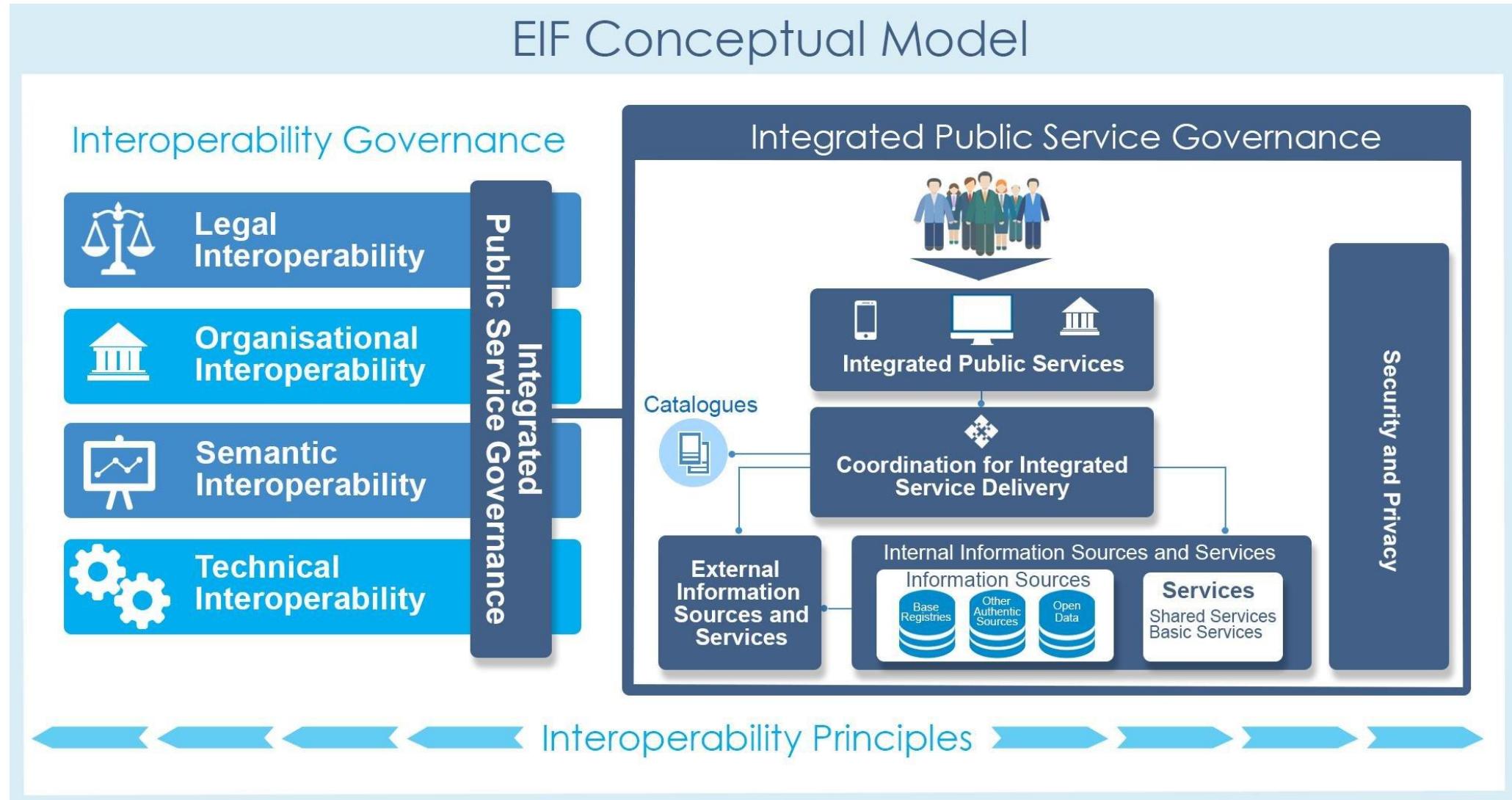
Физический уровень

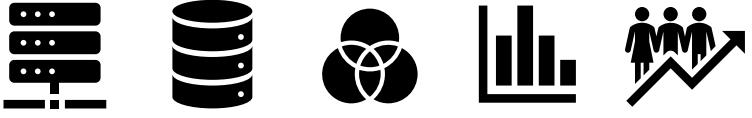
# Интероперабельность – вариант описания

- С 2017 года имеется очень неплохое описание практик интероперабельности, подготовленное *NIFO – National Interoperability Framework Observatory* (<http://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory>)
  - Часть инициативы *JoinUp* с лозунгом *“Share and reuse Interoperability solutions for public administrations, businesses and citizens”*
- *European Interoperability Framework (EIF)* (<http://ec.europa.eu/isa2/eif>)
  - Обзор: *Sharma G. eIDAS and the new European Interoperability Framework – One step closer to the Single Market. 2018* (<http://www.cryptomathic.com/news-events/blog/eidas-and-the-new-european-interoperability-framework-one-step-closer-to-the-single-market>)



# Интероперабельность – EIF





## Качество данных – введение

- ✓ Три главных аспекта
- ✓ Характеристики качества данных
- ✓ Стандартизация

# Оценка качества данных

- Три кита:
  1. **Полнота** [*completeness*]
    - Иногда очень тяжело проверить, особенно при работе с подмножествами узлов социальных сетей
  2. **Корректность** [*correctness*] и **целостность** [*integrity*]
    - Задача распадается на две:
      - Установление критериев корректности
      - Проверка критериев
  3. **Актуальность** [*actuality*]
    - Здесь всё проще, но иногда тяжело определить валидность отметок времени
      - Кто сталкивался с присвоением отметок времени «от балды»?
- *Open Data Institute Certificates* (<https://certificates.theodi.org>)
- *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era* (<http://datascience.codata.org/articles/10.5334/dsj-2015-002>) – следующий слайд

# Характеристики качества источников

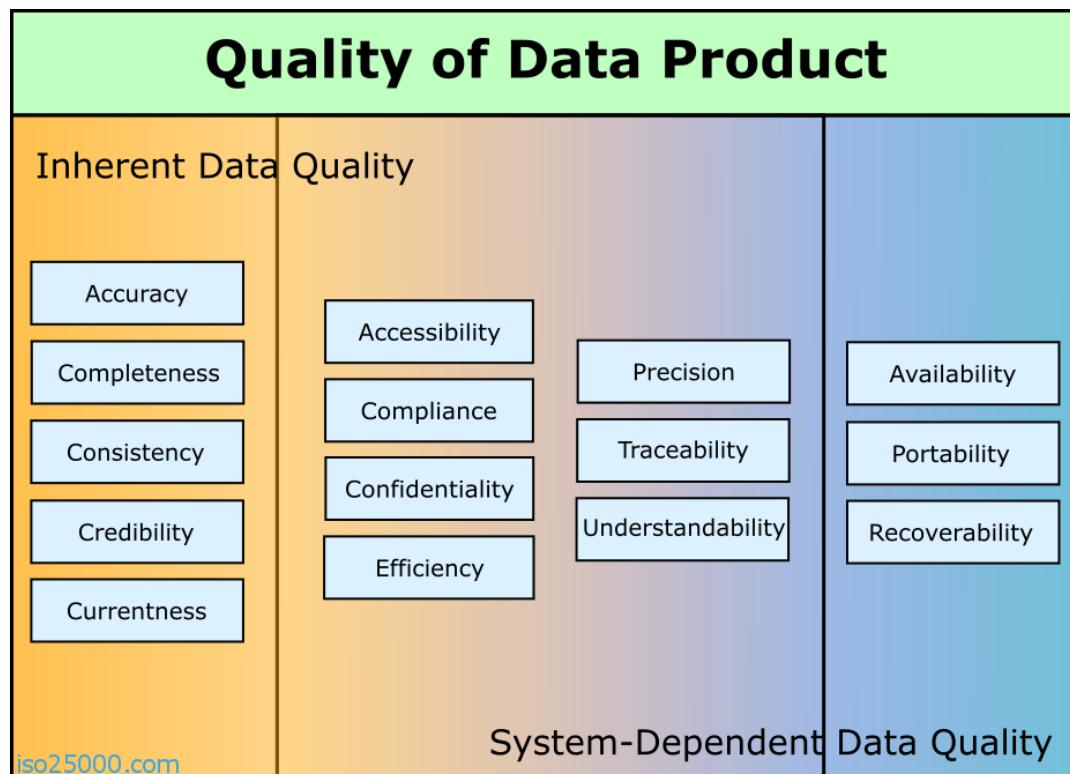


# Стандарты группы ISO 8000

- Группа стандартов **ISO/TS 8000** (<http://www.iso.org/committee/54158/x/catalogue/>)
  - Более 20 частей! Среди них:
    - ISO 8000-1:2022 “*Data quality. Part 1: Overview*”
    - ISO 8000-2:2020 “*Data quality. Part 2: Vocabulary*”
    - ISO 8000-8:2015 “*Data quality. Part 8: Information and data quality: Concepts and measuring*”
  - Перевод:
    - ГОСТ Р 56214-2014/ISO/TS 8000-1:2011 Качество данных. Часть 1. Обзор (<http://docs.cntd.ru/document/1200114769>) – **ещё не обновили!**
    - ГОСТ Р ИСО 8000-2-2019 «Качество данных. Часть 2. Словарь» (<https://docs.cntd.ru/document/1200169126>)
  - В основе стандартов комплекса 8000 лежат следующие принципы:
    - a) качество данных затрагивает данные, имеющие определенное назначение, т.е. участвующие в принятии какого-либо решения
    - b) качество данных затрагивает нужные и подходящие данные, уместные в подходящем месте в подходящее время
    - c) качество данных отвечает требованиям потребителя
    - d) качество данных предотвращает повторение дефектов данных и сокращает избыточные расходы

# SQuaRE и продукт данных

- Группа ISO/IEC 25000, больше известная как **SQuaRE** (*System and Software Quality Requirements and Evaluation*) вводит понятие «продукта данных»
  - ISO 25000 PORTAL (<http://iso25000.com>)



# Управление данными в области ИИ

- ISO/IEC 8183:2023 Artificial intelligence — Data life cycle framework (<http://www.iso.org/standard/83002.html>)
  - This document defines the stages and identifies associated actions for data processing throughout the artificial intelligence (AI) system life cycle, including acquisition, creation, development, deployment, maintenance and decommissioning. This document does not define specific services, platforms or tools. This document is applicable to all organizations, regardless of type, size or nature, that use data in the development and use of AI systems.
- ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) (<https://www.iso.org/standard/74438.html>)
- ISO/IEC 5259-1:2024 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples (<http://www.iso.org/standard/81088.html>)
- ISO/IEC 5259-4:2024 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework (<http://www.iso.org/standard/81093.html>)



## Формализация знания

- ✓ Знания явные и неявные
- ✓ Измерения и заземление
- ✓ Формальные онтологии
- ✓ Онтологическое моделирование
  - ✓ Логики и логический вывод
  - ✓ Графы знаний и семантические сети
- ✓ Онтологизация метаданных

# Явные и неявные знания

- Знание:
  - **Явное**, эксплицитное [*explicit*] – только им можно п полноценно управлять!
  - **Неявное**, имплицитное [*implicit*]
  - **Практическое**, тацитное [*tacit*]
- Определений и дистинкций очень много!
- Например, в *Alexander R. Implicit, Tacit, or Explicit: All Knowledge Is Valuable.*  
(<http://bloomfire.com/blog/implicit-tacit-explicit-knowledge/>)
  - **Explicit Knowledge:** *Knowledge that is easy to articulate, write down, and share*
  - **Implicit Knowledge:** *The application of explicit knowledge*
    - *Skills that are transferable from one job to another are one example of implicit knowledge*
  - **Tacit Knowledge:** *Knowledge gained from personal experience that is more difficult to express*

# Реальный мир и актуальные знания

- Измерение и контекст измерения
- Онтологизация измерения в различных парадигмах
- Шкалы и единицы измерения – основы онтологизации на практике
- Заземления различных видов и 4D/ND-физически-основанные пространственно-временные отношения
- Объекто-признаковые данные и формальные понятия
  - См. прошлогоднее выступление  
([http://github.com/agirussia/agirussia.github.io/blob/main/presentations/2024/Neznanov-ObjectAttributeData\\_Part1\\_2024-05.pdf](http://github.com/agirussia/agirussia.github.io/blob/main/presentations/2024/Neznanov-ObjectAttributeData_Part1_2024-05.pdf))

# Main ISO standard for quantities and units

- ISO 80000-1:2022 *Quantities and units — Part 1: General*  
(<http://www.iso.org/standard/76921.html>)
- ISO 80000-2:2019 *Quantities and units — Part 2: Mathematics*  
(<http://www.iso.org/standard/64973.html>)
- ISO 80000-3:2019 *Quantities and units — Part 3: Space and time*  
(<http://www.iso.org/standard/64974.html>)
- ISO 80000-4:2019 *Quantities and units — Part 4: Mechanics*  
(<http://www.iso.org/standard/64975.html>)
- ...
- ISO 80000-12:2019 *Quantities and units — Part 12: Condensed matter physics*  
(<http://www.iso.org/standard/63480.html>)
- IEC 80000-13:2008 *Quantities and units — Part 13: Information science and technology* (<http://www.iso.org/standard/31898.html>)

# Формальные онтологии (1)

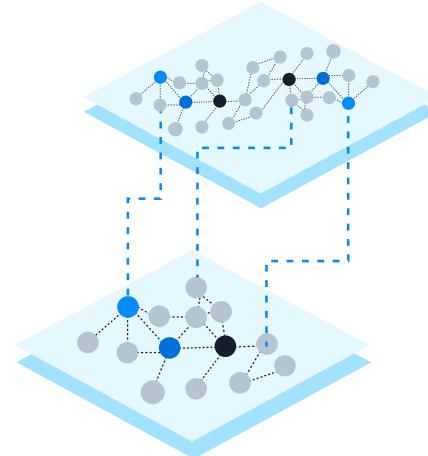
- Термин «онтология» используется в нескольких различных значениях
  1. Наука, изучающая характеристики бытия и сущности
    - Так определена ещё Аристотелем – философская дисциплина
  2. Средство формального моделирования знаний о некоторой системе или предметной области (в самом широком смысле) – формальная онтология
    - Чаще всего используемое в наше время техническими специалистами
  3. База формальных (эксплицитных) знаний – инженерная онтология
    - Используется специалистами в ИТ для различия с базами знаний в широком смысле
- 1993 г. – Т. Грубер определил формальную онтологию как **эксплицитную спецификацию концептуализации**
  - *Explicit specification of a shared conceptualization!*
  - Это определение является «лингвистически красивым» и наиболее цитируемым

# Формальные онтологии (2)

- 1997 г. – В. Борст определила онтологию как **формальную спецификацию общей концептуализации**, с дополнительным требованием, чтобы данная концептуализация выражала **общее мнение некоторой группы людей**
  - *Borst W.N., Borst W.N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Enschede: Centre for Telematics and Information Technology (CTIT), 1997.*
- 1998 г. – Р. Студер уточнил прагматику термина, предложив понимать онтологию как **формальную, точную спецификацию общей концептуализации**, которая может использоваться **как человеком, так и компьютером**
  - *An ontology is a formal, explicit specification of a shared conceptualization!*
  - Теперь базовый учебник: *Staab S., Studer R. Handbook on Ontologies. Springer, 2009.*

# Онтологическое моделирование – зачем?

- Кратко:
  - Для рассказа компьютеру и самому себе (человеку) о предметной области
- Развёрнуто:
  - Для уточнения понимания
  - Для формализации знаний – онтология по определению **эксплицитна!**
    - Это позволяет создавать «самоописываемые метаданные»
  - Для верификации знания
  - Для обнаружения лакун в знаниях
  - Для обмена знаниями между людьми и информационными системами
  - Для обеспечения фактами процедур логического вывода
  - Для выявления знаний более высокого уровня из массивов низкоуровневого знания
- Онтологии – основа баз знаний
  - И полноценной интероперабельности информационных систем!



# Требования к онтологиям

- Требования Виктора Раскина и Сергея Ниренбурга
  - **Ясность.** Онтологии должны быть понятными и адекватно передавать смысл объектов и отношений между ними
  - **Возможность расширения.** Онтология должна предусматривать возможность добавления новых элементов, без обязательной корректировки уже существующих
  - **Согласованность.** Утверждения, содержащиеся в онтологии, не должны противоречить друг другу, иерархии понятий и связывающим их отношениям
  - **Минимизация избыточной специализации.** Онтология не должна жёстко привязываться к конкретной задаче – желательно обеспечить возможности её использования в будущих задачах
- *Nirenburg S., Raskin V. Ontological Semantic. MIT Press, 2003.*  
[\(http://mitpress.mit.edu/9780262538336/ontological-semantics/\)](http://mitpress.mit.edu/9780262538336/ontological-semantics/)

# Математические основы онтологического моделирования

- Основные способы описания онтологии:
  - Логики различных типов [*logic*]
  - Семантические сети общего вида [*semantic network*]
  - Системы триад (триплетов) [*triplet*]
  - Понятийные структуры [*conceptual structure*]
  - Фреймы [*frame*]
  - ...
- В зависимости от базового математического формализма совершенно по разному могут определяться языки запросов к онтологиям
  - Например, известный язык *SPARQL* работает с онтологиями, использующими в основе язык *RDF*, то есть триплеты
- Исключительно важная веха – стандартизация в 2007 г. базового логического инструментария в виде *ISO/IEC 24707:2018 Common Logic* (<http://iso-commonlogic.org>)
  - Позволяет использовать три различных нотации (включая *XML*) для работы с понятийными структурами и логикой предикатов первого порядка (с некоторыми расширениями)
  - Стандарт сложен для повседневности, но недостаточен для онтологий

# Значимость логики – пример

- 2011 – по сумме двух игр *IBM Watson* заработал 77 тысяч долларов, **более чем втрое** обогнав сильнейших участников *Jeopardy* – Кена Дженнингса (*Ken Jennings*) и Брэда Руттера (*Brad Rutter*)
  - *Jeopardy* – американский аналог «Своей игры»
- Одна из логических ошибок Watson'a:
  - Простой вопрос «назовите аэропорт в Америке, названный в честь героя 2-й мировой войны» *Watson* не смог назвать О'Хара (Чикаго), хотя у него было знание имени этого аэропорта и знание того, что О'Хара получил высший орден США за Боевые Заслуги, но **не было заложено знание** о том, что: «получатель Ордена за Боевые Заслуги» есть «герой»

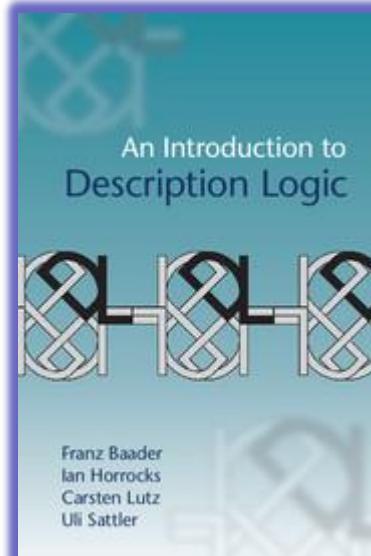
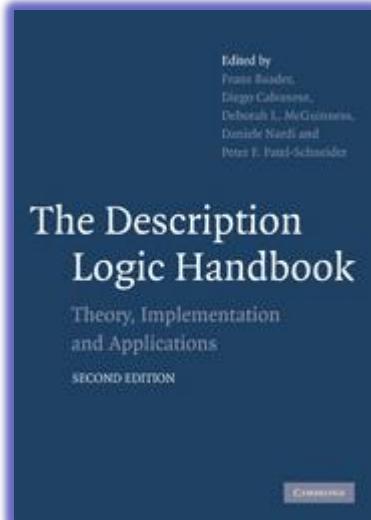


# Дескриптивная/дескрипционная логика

- Минимально необходимая логика для вывода в онтологиях на основе семантических отношений!
- Дескриптивная/дескрипционная логика [*descriptive/description logic*]
  - *MLWiki – Descriptive Logic* ([http://mlwiki.org/index.php/Descriptive\\_Logic](http://mlwiki.org/index.php/Descriptive_Logic))
  - *Krötzsch M., Simancik F., Horrocks I. A Description Logic Primer* (<http://arxiv.org/abs/1201.4089>)
  - *Horrocks I. Description Logic: A Formal Foundation for Ontology Languages and Tools* ([http://www.cs.ox.ac.uk/ian.horrocks/Seminars/download/Horrocks\\_Ian\\_pt1.pdf](http://www.cs.ox.ac.uk/ian.horrocks/Seminars/download/Horrocks_Ian_pt1.pdf))
- Решатели (подсистемы логического вывода):
  - *OWL@Manchester – List of Reasoners* (<http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/>)
  - *Complexity of reasoning in Description Logics* (<http://www.cs.man.ac.uk/~ezolin/dl/>)

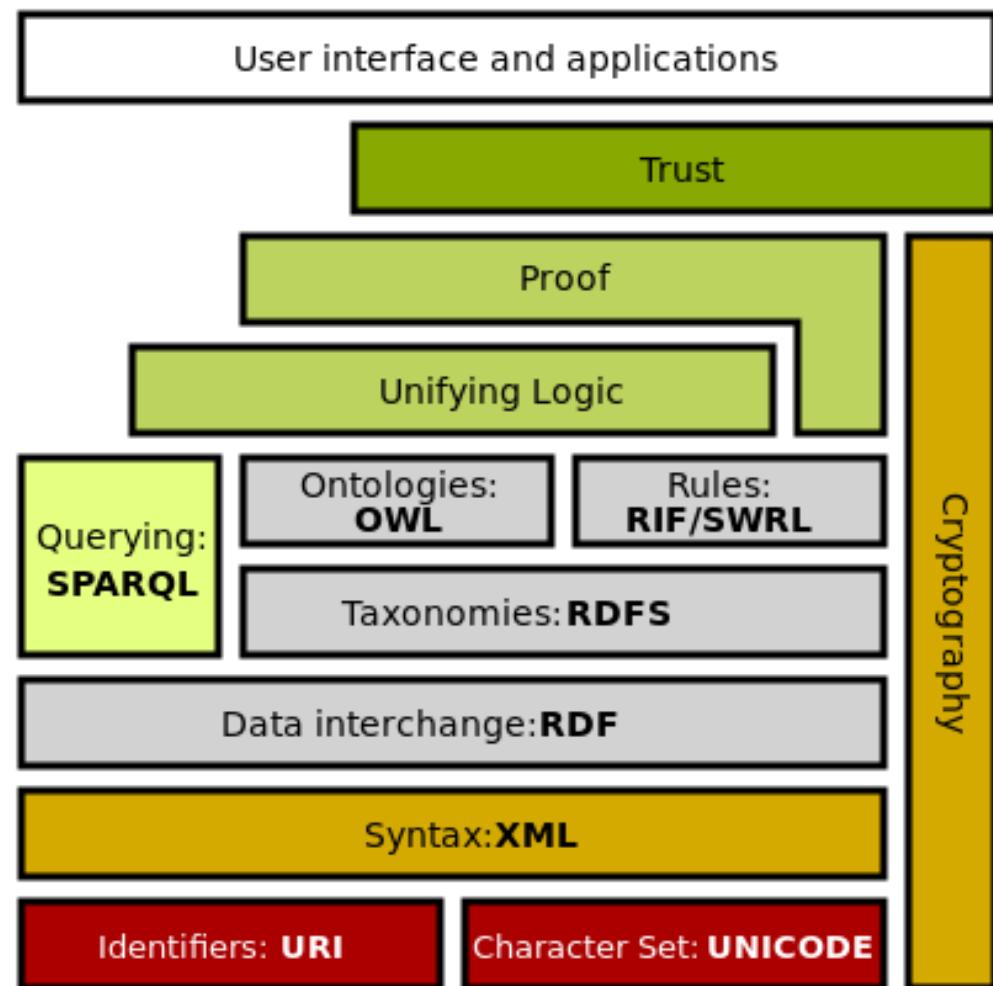
# DL – Классические альманахи и учебник

- *The Description Logic Handbook: Theory, Implementation and Applications, 2<sup>nd</sup> Ed. Cambridge University Press, 2007.* 624 p.
- *The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2003.* 574 p.
- *Baader F., Horrocks I., Lutz C., Sattler U. An Introduction to Description Logic. Cambridge University Press, 2017.* 264 p.



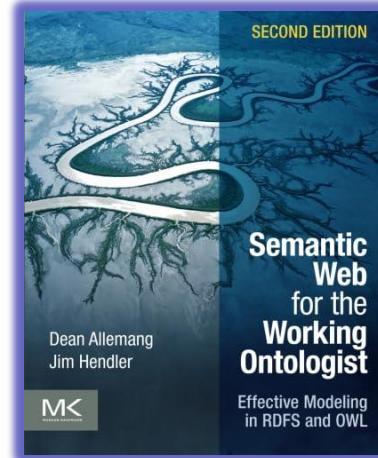
# Иллюстрация одного из стеков

- Стандартный стек для Semantic Web (<http://www.w3.org/standards/semanticweb/>):
  - XML (<http://www.w3.org/XML/>)
    - From 1996
  - RDF (<http://www.w3.org/RDF/>)
    - From 1999
  - OWL (<http://www.w3.org/OWL/>)
    - From 2004
  - SPARQL (<http://www.w3.org/TR/sparql11-overview/>)
    - From 2008
  - SKOS (<http://www.w3.org/TR/skos-reference/>)
    - From 2005
  - +++

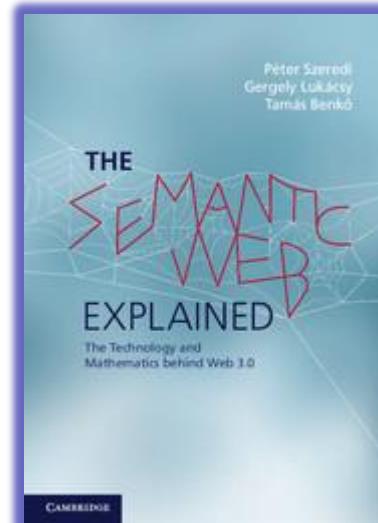


# *Semantic Web – Хорошее описание*

- *Allemang D., Hendler J. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. 2<sup>nd</sup> Ed. Morgan Kaufmann, 2011.* 384 p.



- *Szeredi P., Lukácsy G., Benkő T. The Semantic Web Explained. The Technology and Mathematics behind Web 3.0. Cambridge University Press, 2014.* 478 p.



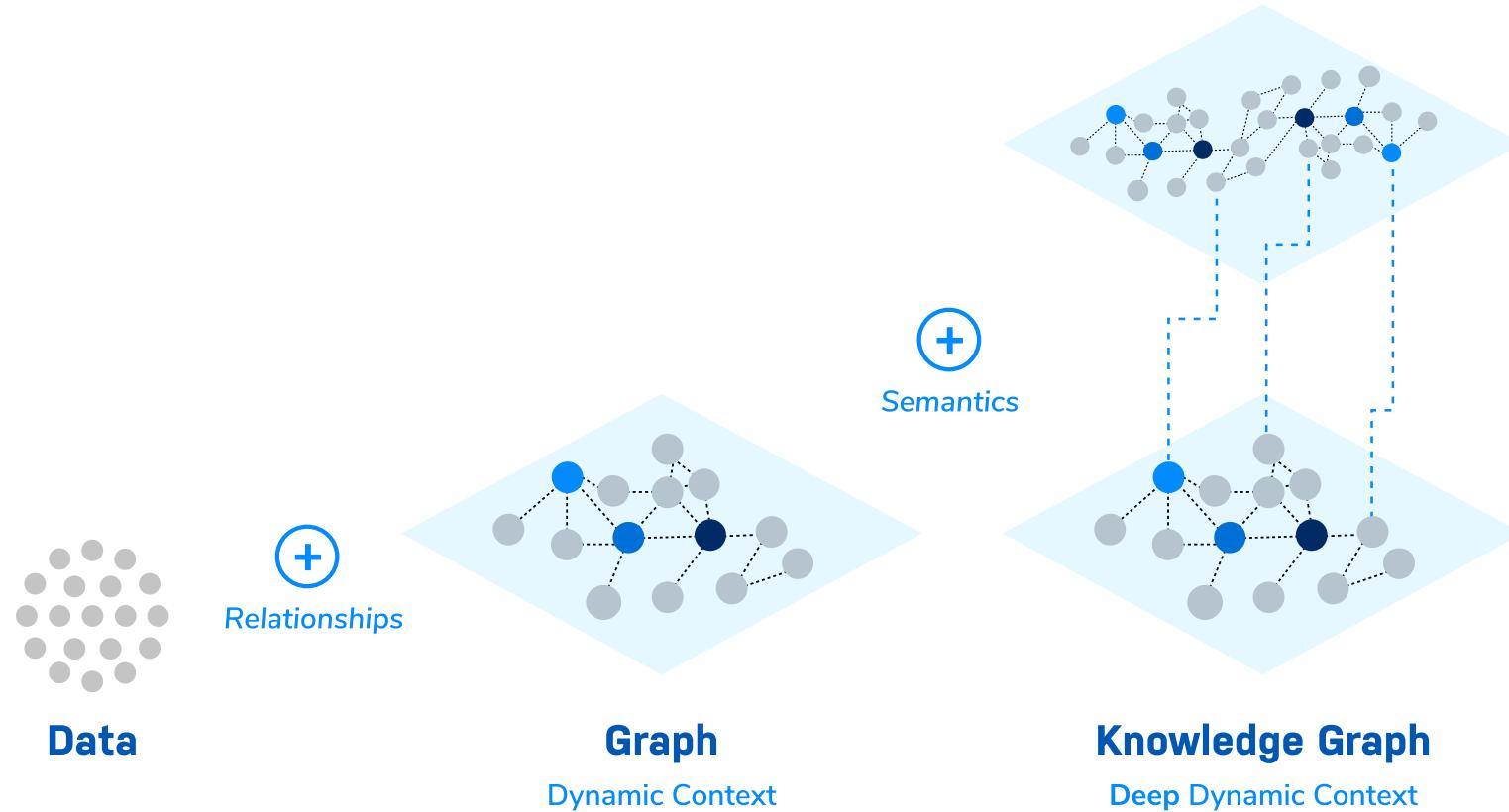
# *Semantic Web Standards with Versioning Hell*

- W3C standards and drafts – DATA ([http://www.w3.org/TR/?tags\[0\]=data](http://www.w3.org/TR/?tags[0]=data))
  - XML - Extensible Markup Language (XML) 1.1 (<http://www.w3.org/TR/xml11/>)
    - W3C Recommendation **2006-08-16**
  - XQuery – XQuery 3.0: An XML Query Language (<http://www.w3.org/TR/xquery-30/>)
    - W3C Recommendation **2014-04-08**
- RDF – RDF 1.2 Concepts and Abstract Syntax (<http://www.w3.org/TR/rdf12-concepts/>)
  - W3C Working Draft **2024-04-16!**
    - Сразу видно, что важно развивать...
    - RDF 1.1 Semantics (<http://www.w3.org/TR/rdf11-mt/>), W3C Recommendation **2014-02-25**
- OWL – OWL 2 Web Ontology Language (<http://www.w3.org/TR/owl2-overview/>)
  - W3C Recommendation **2012-12-11**
- SHACL – Shapes Constraint Language (SHACL) (<http://www.w3.org/TR/shacl/>)
  - W3C Recommendation **2017-07-20**
    - Knublauch H. SHACL and OWL Compared. 2017 (<http://spinrdf.org/shacl-and-owl.html>)

# Семантические сети и графы знаний

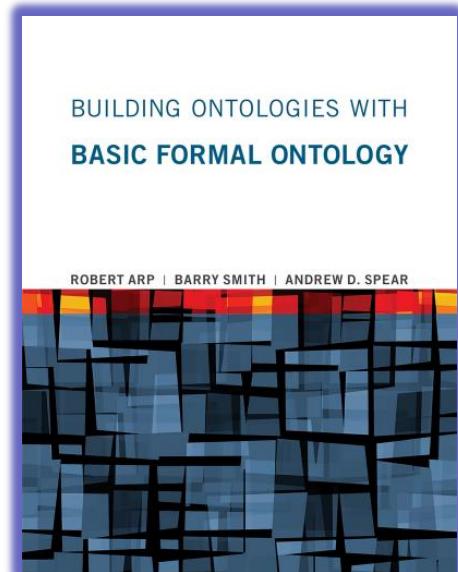
- Формализация, интерпретация, визуализация, хранение

- *Neo4J WP – Knowledge Graphs* (<http://neo4j.com/use-cases/knowledge-graph/>)

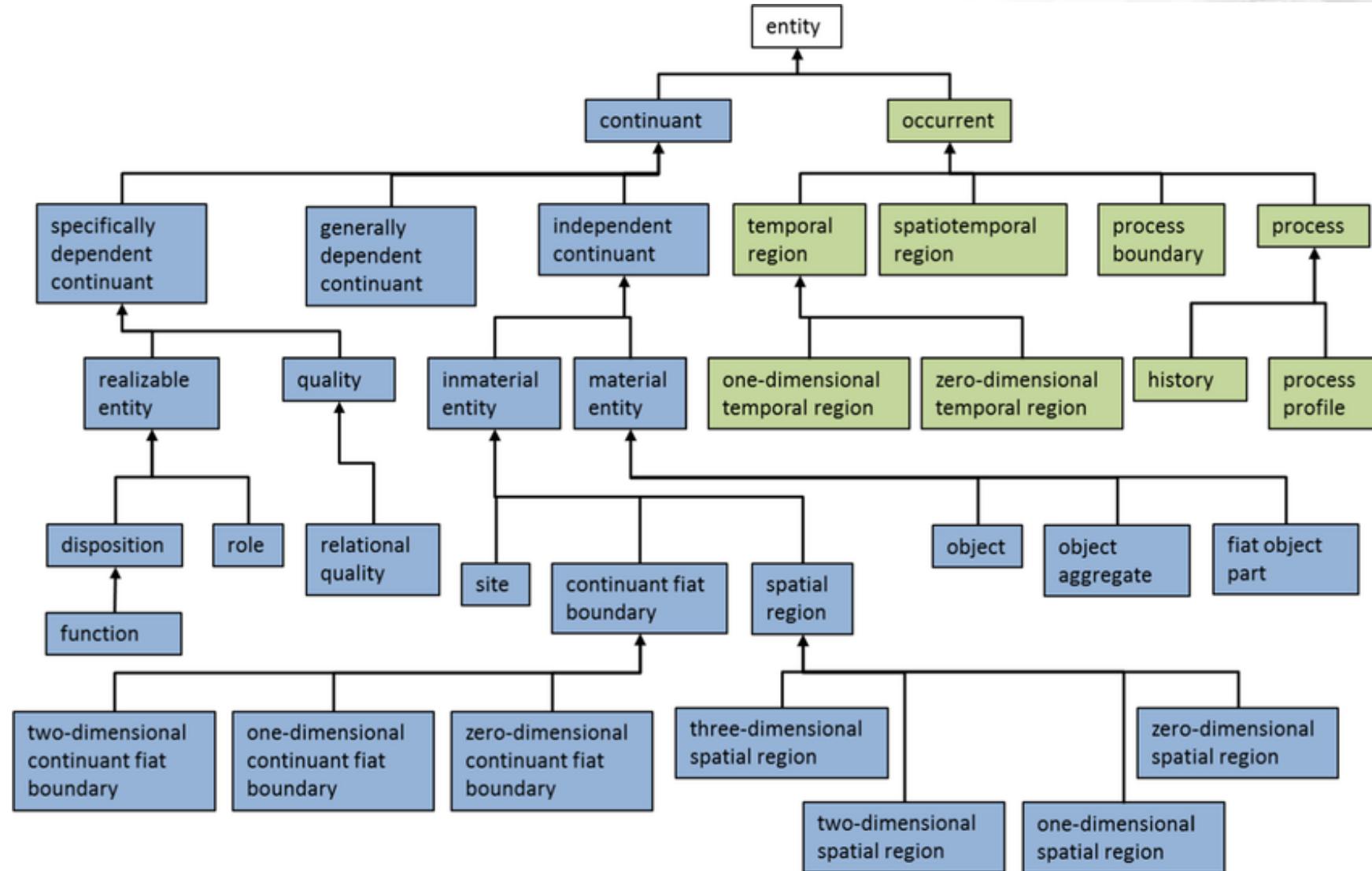


# Онтологии верхнего уровня (*TLO*)

- ISO/IEC 21838-2:2021 *Information technology – Top-level ontologies (TLO)*  
*Part 1: Requirements* (<http://www.iso.org/standard/71954.html>)
- ISO/IEC 21838-2:2021 *Information technology – Top-level ontologies (TLO)*  
*Part 2: Basic Formal Ontology (BFO)* (<http://www.iso.org/standard/74572.html>)
- *BFO – Basic Formal Ontology* (<http://basic-formal-ontology.org>)
  - “a small, upper-level ontology that is designed for use in supporting information retrieval, analysis and integration in scientific and other domains”
    - *Basic Formal Ontology 2.0 – Specification and User’s Guide* (<http://github.com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference.pdf>)
    - *Arp R., Smith B., Spear A.D. Building Ontologies with Basic Formal Ontology. The MIT Press, 2015. 220 p.*



# Основные сущности BFO

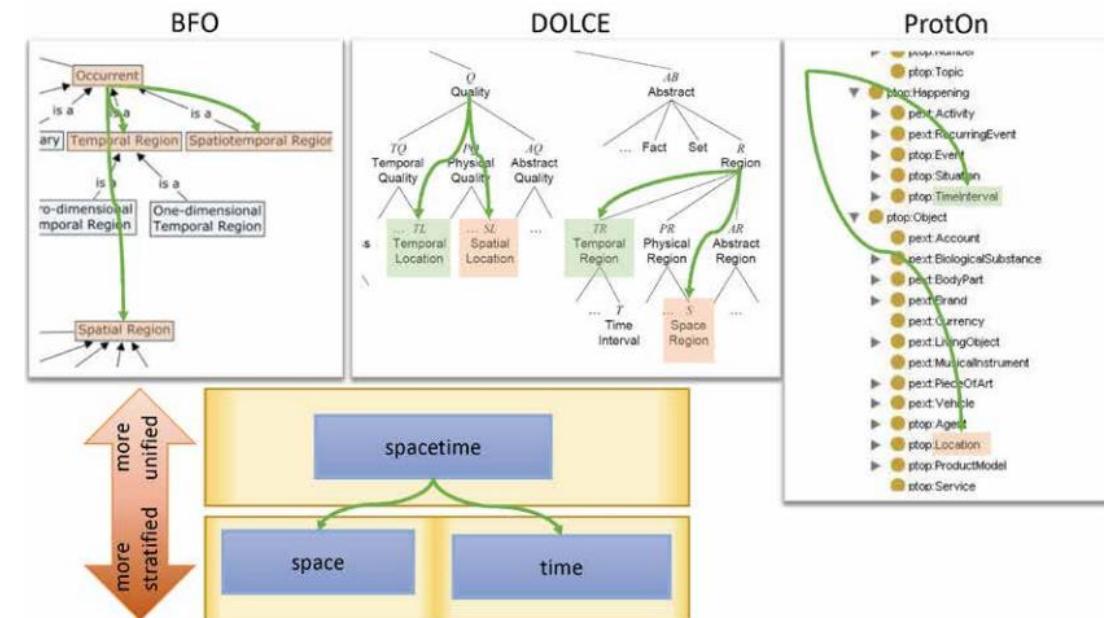
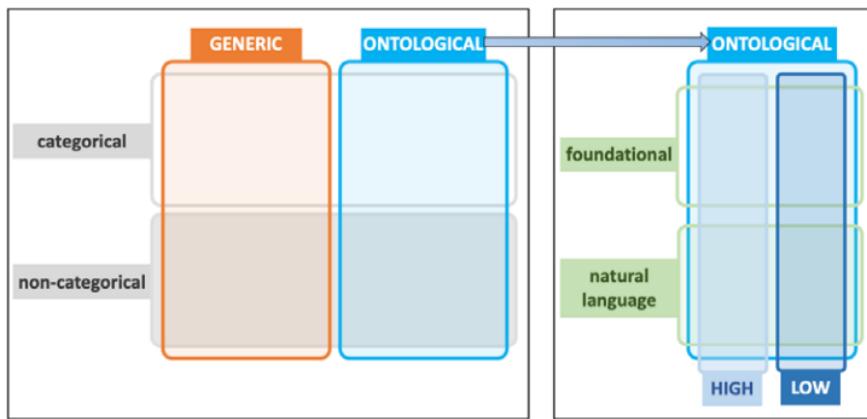


# Обзор других TLO

- Partridge C., Mitchell A., et al. *A Survey of Top-Level Ontologies*. 2020 ([http://www.cdbb.cam.ac.uk/files/a\\_survey\\_of\\_top-level\\_ontologies\\_lowres.pdf](http://www.cdbb.cam.ac.uk/files/a_survey_of_top-level_ontologies_lowres.pdf))



- General classification of the TLOs:

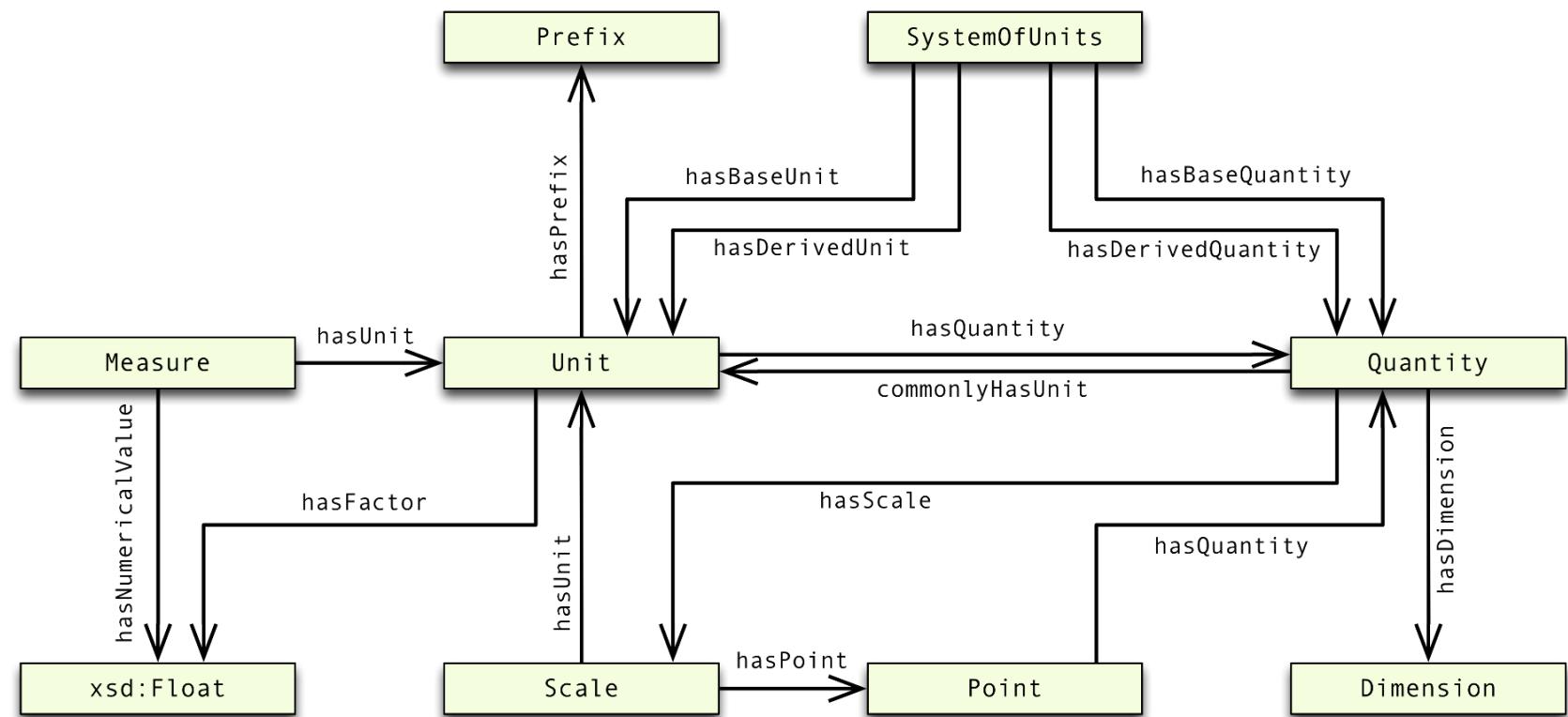


- Separating spacetime – TLO examples:

# Общие онтологии (пример единиц измерения)

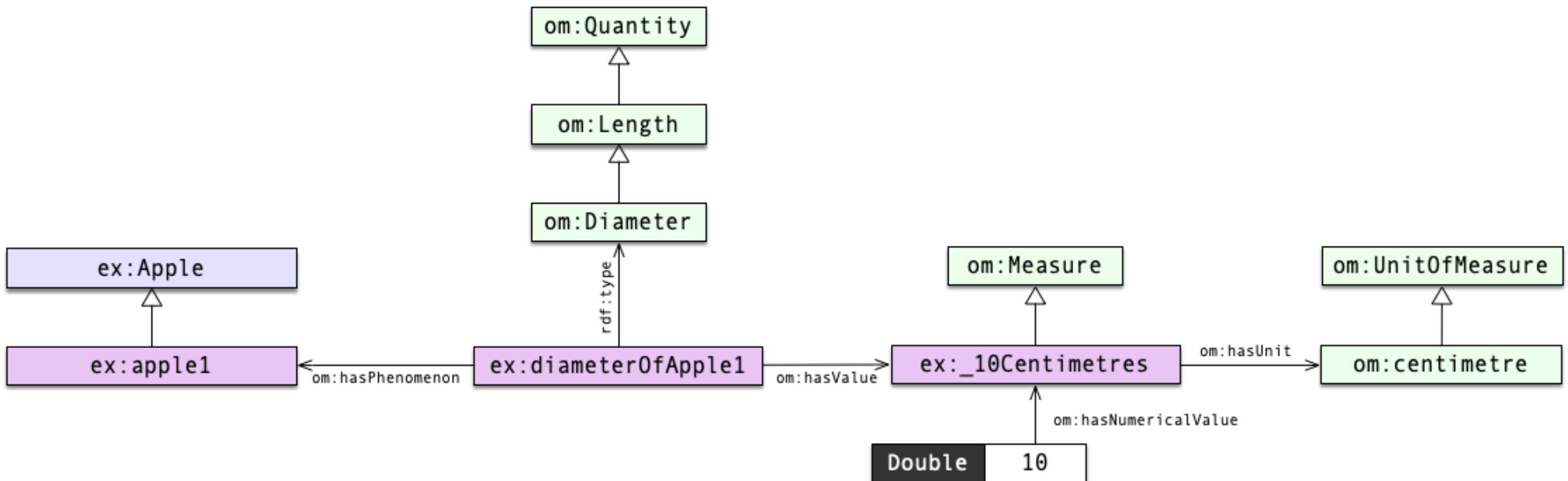
- Онтология единиц измерения [*Ontology of units of Measure*]

- [GitHub] *HajoRijgersberg/OM* (<http://github.com/HajoRijgersberg/OM>)
  - Current version – 2.0

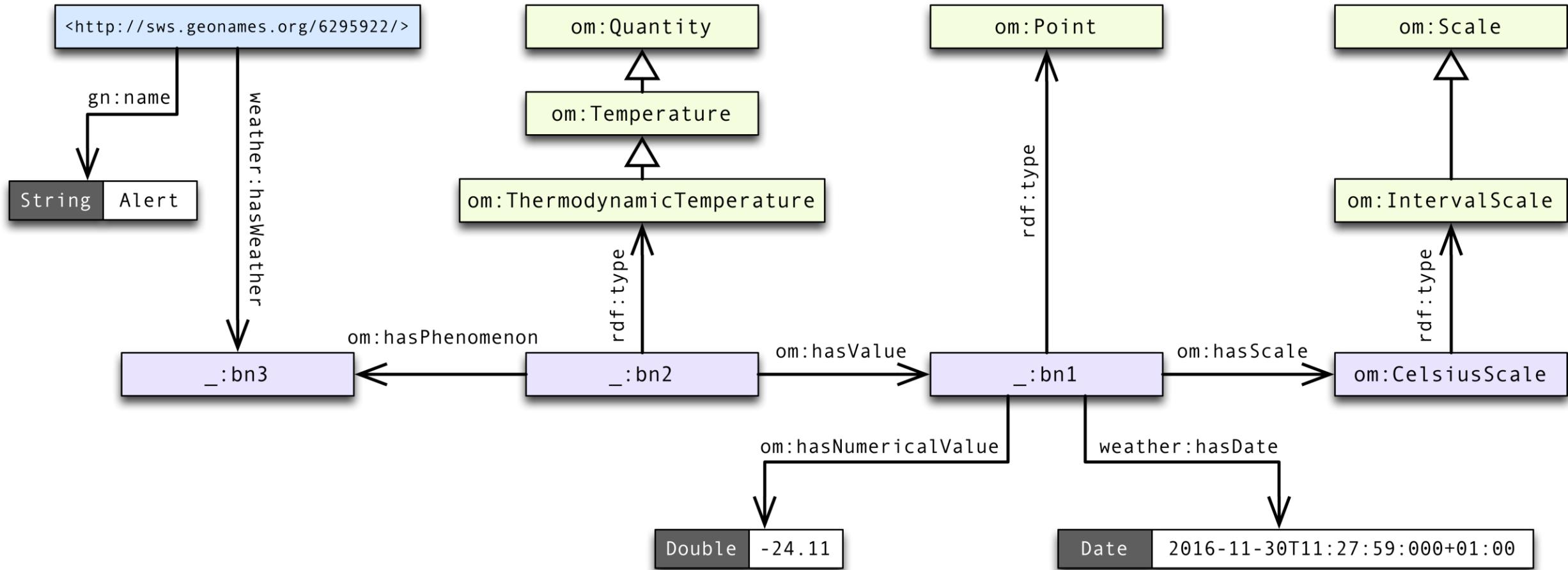


# Пример использования онтологии

- Представление «конкретного акта измерения» с привязкой к:
  - Онтологии единиц измерения («сантиметр», «диаметр»)
  - Онтологии предметной области («яблоко»)



# Пример измерения в контексте (температура)



# Системы управления базами знаний (СУБЗ)

- По аналогии с СУБД выделился класс ПО под названием **системы управления базами знаний** (СУБЗ)
- Но сейчас восторжествовала точка зрения, что любая приличная ИС должна быть основана на знании, то есть быть *Knowledge Based System*
  - В результате компоненты для работы со знанием сейчас обычно называют *Knowledge Management System (KMS)* – системами управления знаниями
  - Они очень разнообразны и поддерживают различные формы представления знания
- Главный академический проект – *Protégé* (<http://protege.stanford.edu>)
  - Tudorache T., Nyulas C., Noy N.F., Musen M.A. WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web, Semantic Web, 4(1), 2013, pp. 89–99 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691821/>)

# Protégé - Entity Classes and Individuals

File Edit View Reasoner Tools Refactor Window Help

om-2 (<http://www.ontology-of-units-of-measure.org/resource/om-2/>) : [F:\OneDrive\Courses\Curs-ED...]

quantity > acceleration > gravitational acceleration

Active ontology Entities Individuals by class DL Query

Annotation properties Datatypes Individuals  
Classes Object properties Data properties

Class hierarchy: gravitational acceleration

Annotations: gravitational acceleration

Annotations:  
 rdfs:label [language: en]  
 gravitational acceleration  
 rdfs:label [language: nl]  
 valversneling  
 'alternative label' [language: en]  
 acceleration of free fall  
 'alternative label' [language: nl]  
 gravitatieveldsterkte  
 'alternative label' [language: nl]  
 zwaarteveldsterkte  
 symbol  
 g

Description: gravitational acceleration

Equivalent To  
 SubClass Of  
 acceleration  
 General class axioms  
 SubClass Of (Anonymous Ancestor)  
 'has dimension' value 'acceleration dimension'  
 'has value' only ('has unit' only 'acceleration unit')  
 Instances  
 Target for Key  
 Disjoint With

No Reasoner set. Select a reasoner from the Reasoner menu  Show Inferences

File Edit View Reasoner Tools Refactor Window Help

om-2 (<http://www.ontology-of-units-of-measure.org/resource/om-2/>) : [F:\OneDrive\Courses\Curs-ED...]

system of units

Active ontology Entities Individuals by class DL Query

Class hierarchy: system of units

Annotations: International System of Units

Annotations:  
 rdfs:label [language: en]  
 International System of Units  
 rdfs:label [language: zh]  
 国际单位制  
 rdfs:label [language: nl]  
 Internationale Stelsel van Eenheden  
 abbreviation  
 SI

Description: International System of Units

Types:  
 'system of units'

Direct instances: International System

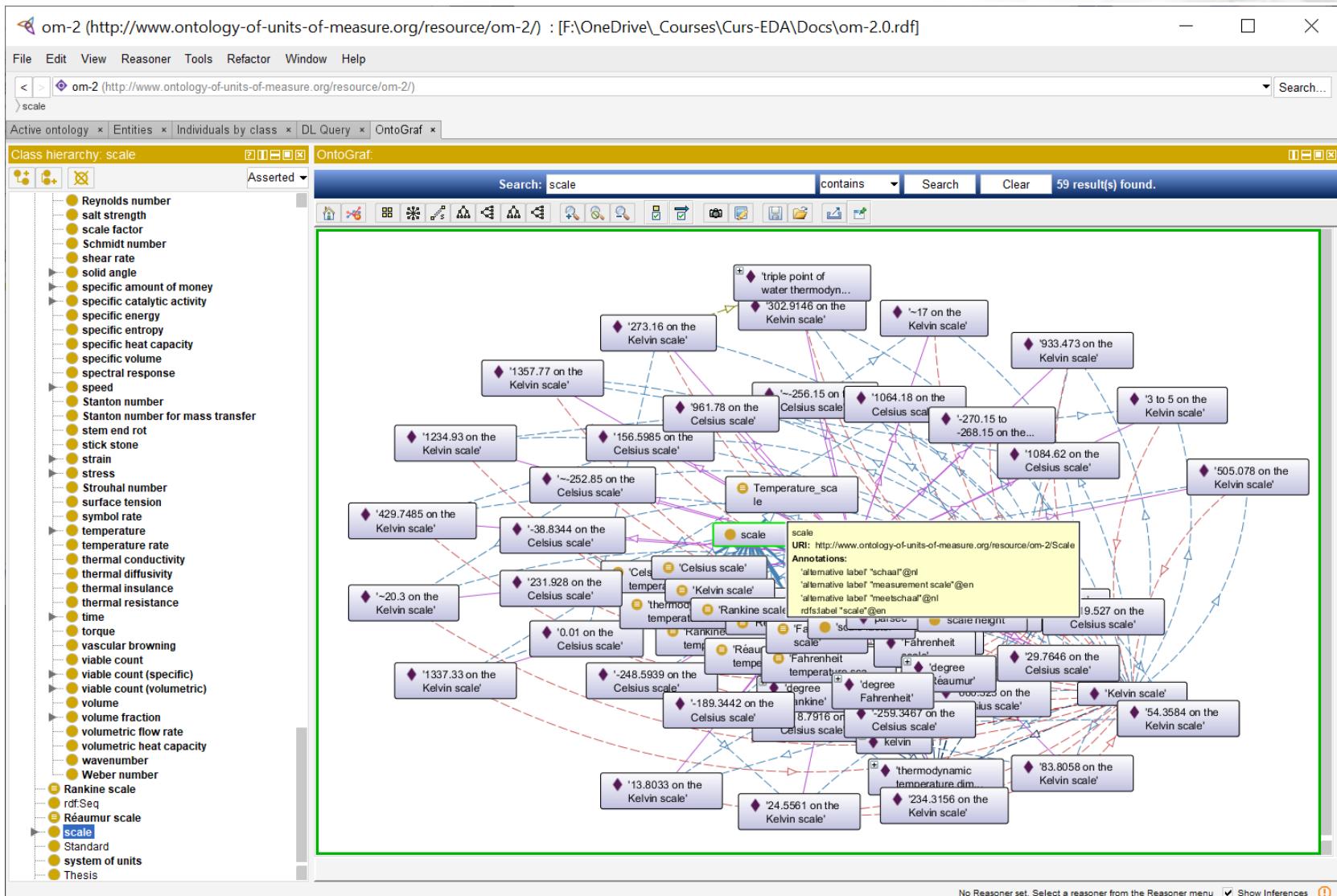
For: 'system of units'  
 centimetre-gram-second electromagnetic system of units  
 centimetre-gram-second electrostatic system of units  
 centimetre-gram-second system of units  
 centimetre-gram-second-biot system of units  
 centimetre-gram-second-franklin system of units  
 Gaussian system of units  
 International System of Units  
 metre-kilogram-second-ampere system of units

Object property assertions:

- 'has derived unit' 'radian per second'
- 'has derived unit' 'zettacandela'
- 'has derived unit' 'femtowebster'
- 'has derived unit' 'petahertz'
- 'has derived unit' 'attoradian'
- 'has derived unit' 'hectometre per second'
- 'has derived unit' 'nanoradian'
- 'has derived unit' 'decahenry'
- 'has derived unit' 'yottakatal'
- 'has derived unit' 'catalytic activity'
- 'has derived unit' 'kilopascal'
- 'has derived unit' 'exawatt'
- 'has derived unit' 'zettametre'
- 'has derived unit' 'zettagram'
- 'has derived unit' 'hectonewton'
- 'has derived unit' 'zeptolumen'
- 'has derived unit' 'yottasecond'
- 'has derived quantity' 'density'
- 'has derived unit' 'attotesla'
- 'has derived unit' 'decakatal'
- 'has derived unit' 'exasiemens'

No Reasoner set. Select a reasoner from the Reasoner menu  Show Inferences

# Protégé - OntoGraf



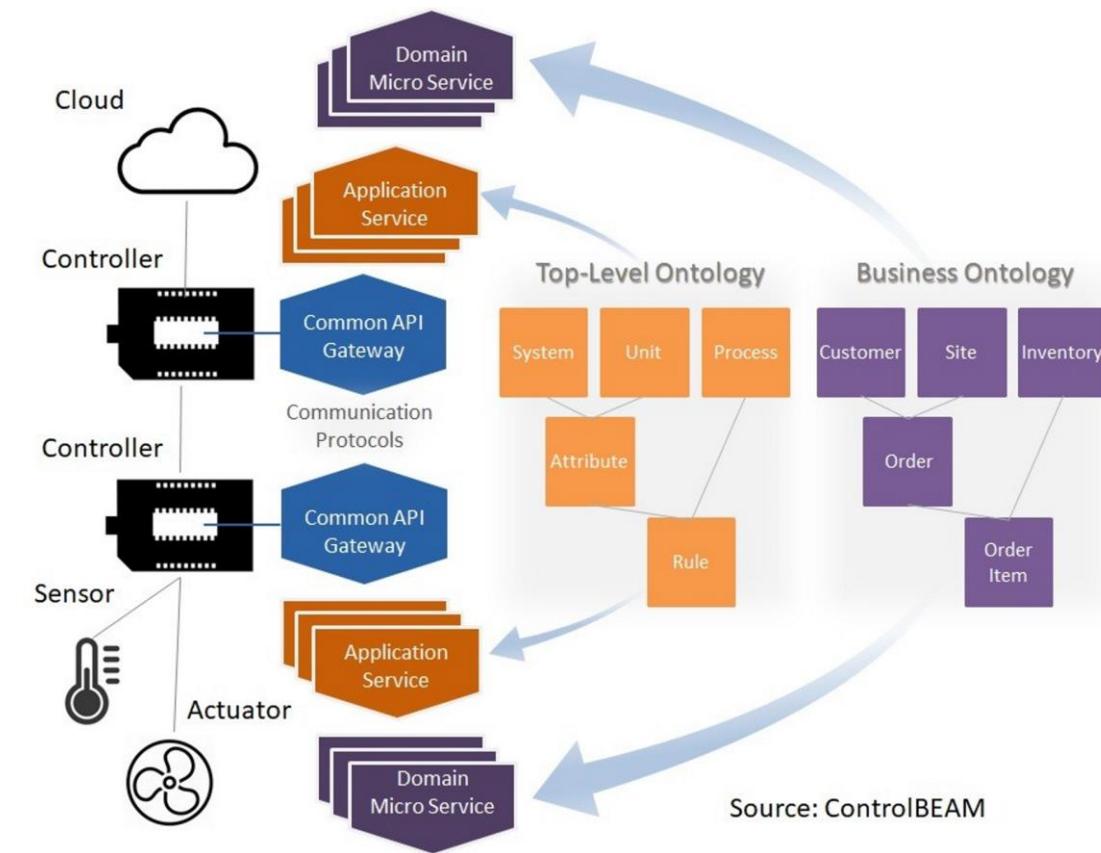
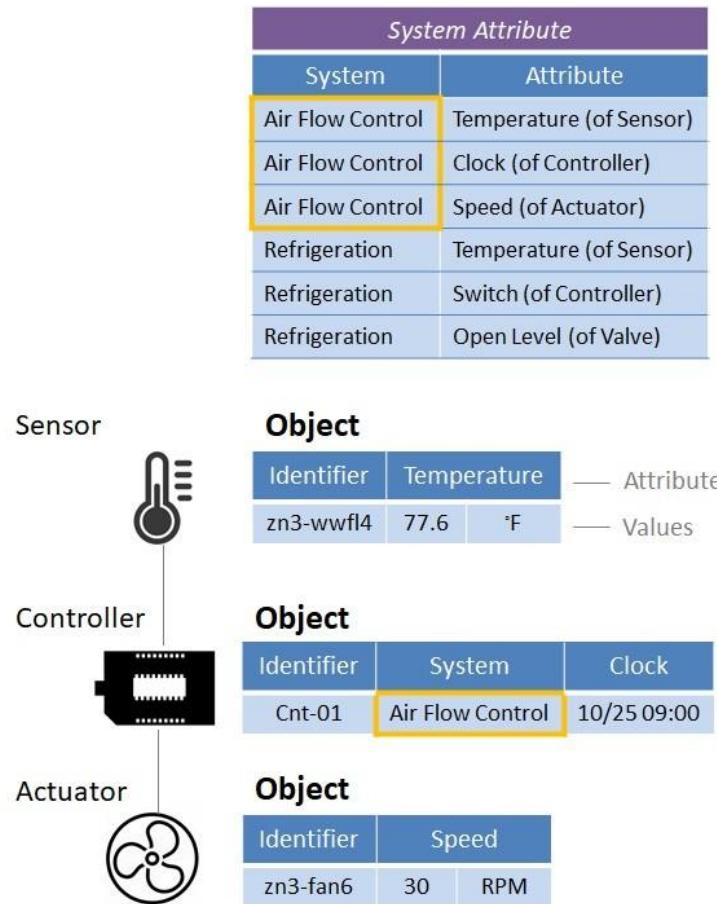


## Онтологизация и БЗ в кровавом интерпрайзе

- ✓ Смысл и специфика
- ✓ СУБЗ
- ✓ Корпоративные платформы управления знаниями
- ✓ Отечественные решения

# Пример: значимость онтологий для IoT-платформ

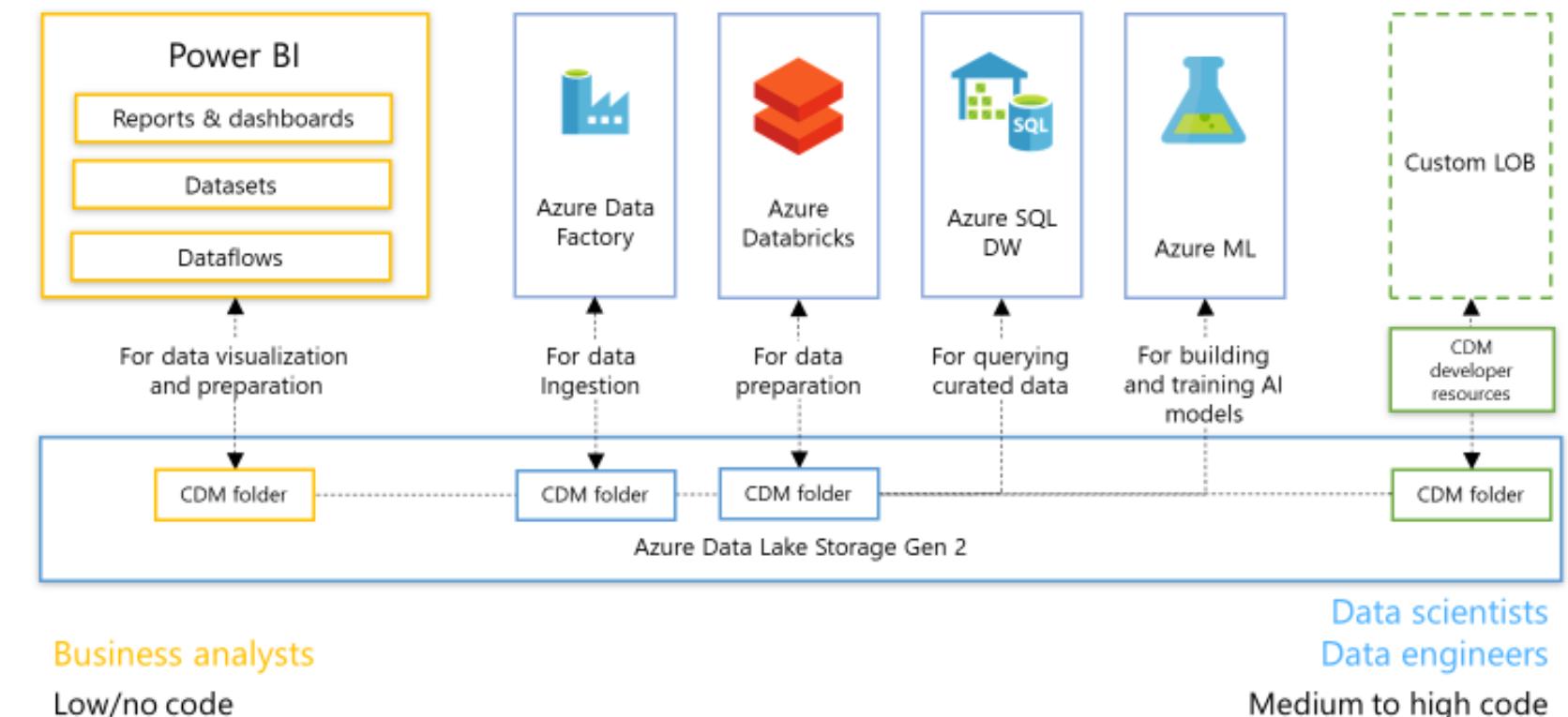
- *Cross-industry semantic interoperability. Part 4: The intersection of business and device ontologies. 2017*  
(<http://embeddedcomputing.com/technology/iot/cross-industry-semantic-interoperability-part-four-the-intersection-of-business-and-device-ontologies>)



# Пример интеграции сервисов через метаданные

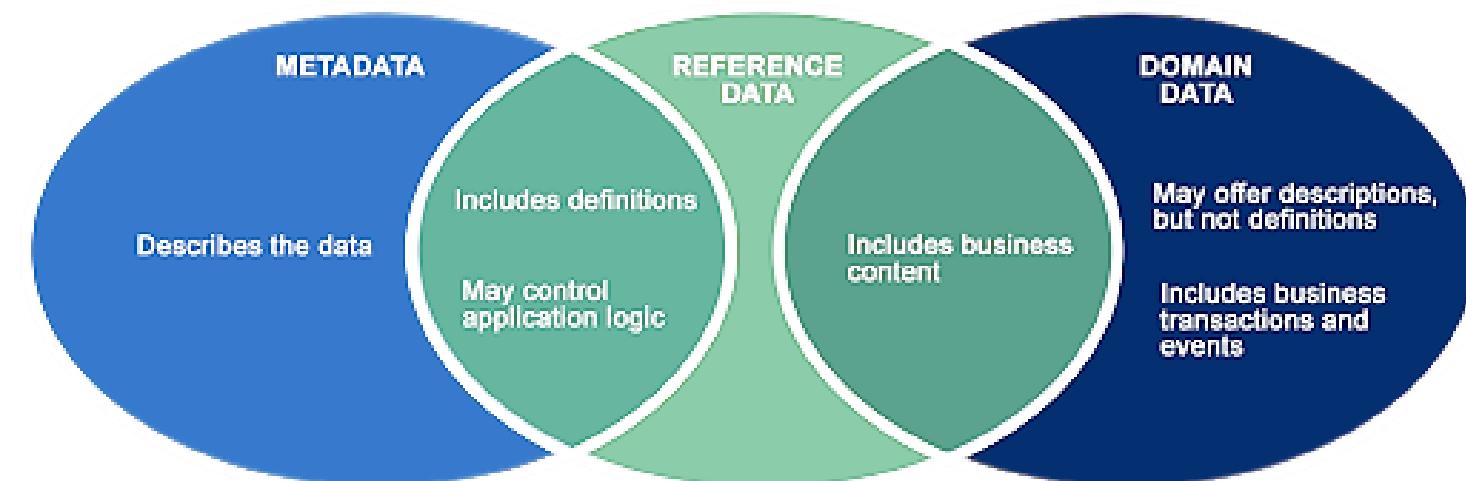
## ● Microsoft – The Common Data Model in Azure Data Lake Storage

- (<https://adatis.co.uk/the-common-data-model-in-azure-data-lake-storage-azure-data-services/>)



# Онтологизация мастер-данных

- Серия статей Алана Фридмана:
  - Time to Rethink Master and Reference Data.* 2019  
(<http://www.linkedin.com/pulse/time-rethink-master-reference-data-alan-freedman/>)
  - Classifying Reference Data for Better Governance.* 2020  
(<http://www.linkedin.com/pulse/classifying-reference-data-better-governance-alan-freedman/>)
  - Your Reference Data Is an Ontology.* 2020  
(<http://www.linkedin.com/pulse/your-reference-data-ontology-alan-freedman>)



# Мастер-данные как знания (пример)

структура мастер-данных



правила валидации

Сущность «Юридическое лицо»
Название: ООО «Омега»
ИНН: -
КПП: 5320015526
ОГРН: 1234567890XXX
Дата образования: 01.01.2012
Дата банкротства: 01.01.2011



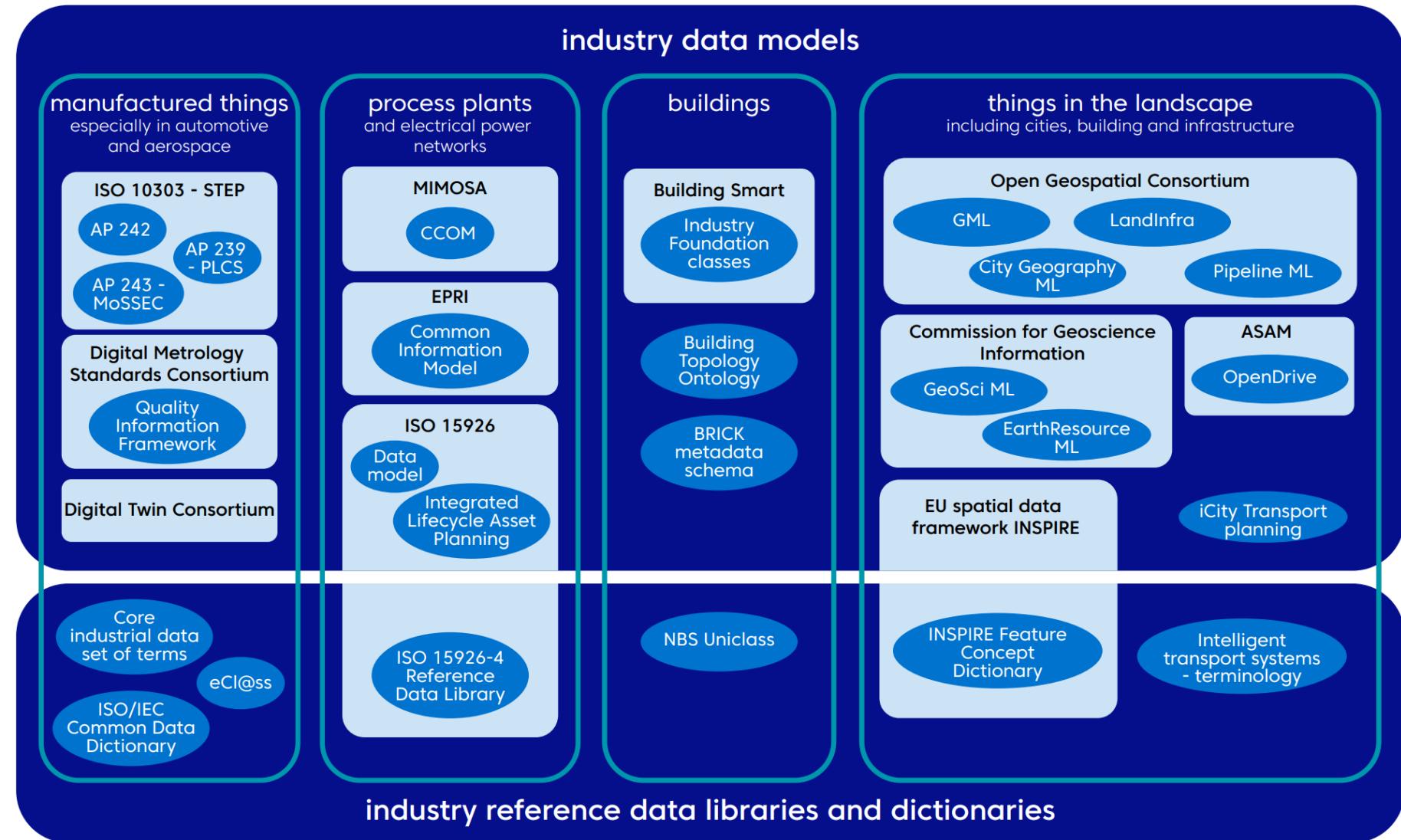
Ошибки валидации
ИНН обязательно должен быть заполнен
КПП должен содержать 9 цифр
ОГРН должен содержать только цифры
Дата банкротства не может быть меньше даты образования

# Обзор онтологизаций моделей данных

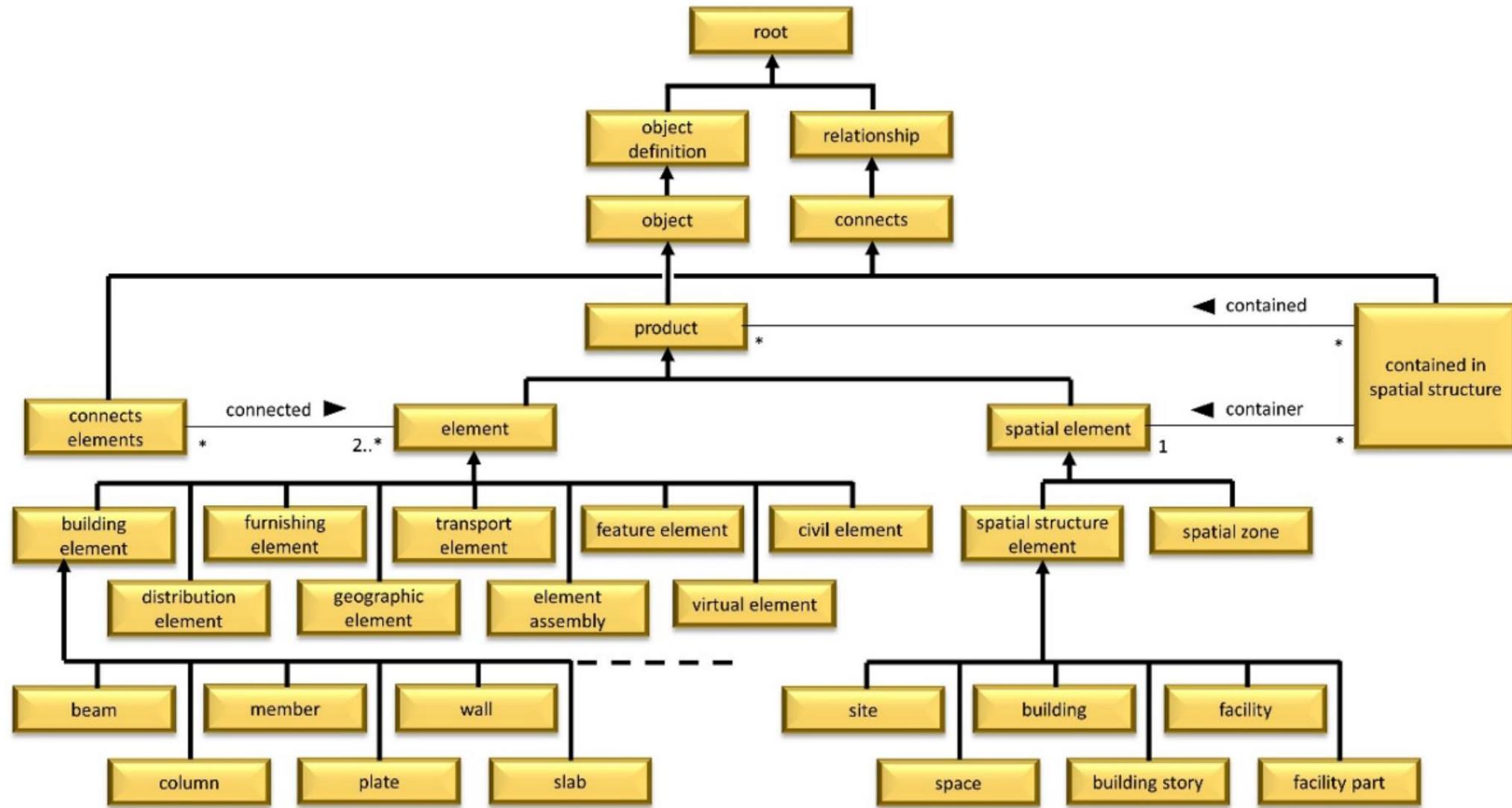
- *Top-Level Ontologies and Industry Data Models Information*
  - *Goal: Management Framework (IMF)*
- *A Survey of Industry Data Models and Reference Data Libraries*  
[\(http://www.cdbb.cam.ac.uk/files/industry\\_data\\_models\\_and\\_reference\\_data\\_libraries\\_0.pdf\)](http://www.cdbb.cam.ac.uk/files/industry_data_models_and_reference_data_libraries_0.pdf)
  - *To identify requirements for, and provide input to, a Foundation Data Model*
  - *Based on ideas of ISO 10303-1:2024 Industrial automation systems and integration — Product data representation and exchange, Part 1: Overview and fundamental principles*



# *Application areas of the industry data models and libraries*



# *Top of the IFC model*

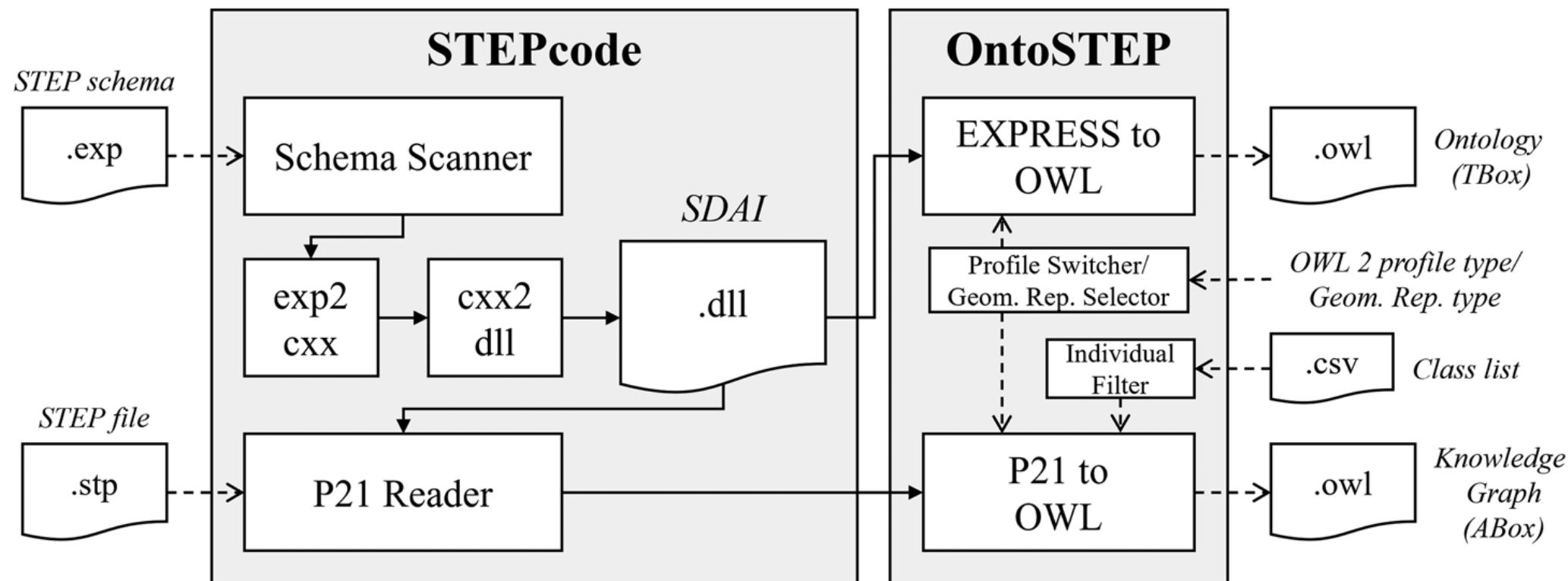


# А обмен данными онтологизирован?

- Да, не только сам ISO 10303!
  - Несколько стандартов с различными парадигмами
- Пример:
  - Krima S., Barbau R., Fiorentini X., Krima S., Sriram R.D., Rachuri S.  
**OntoSTEP: OWL-DL Ontology for STEP (NIST IR 7561)**, NIST US, 2009  
([http://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=901544](http://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=901544))
    - STandard for the Exchange of Product Model Data
  - Kwon S., Monnier L.V., Barbau R., Bernstein, W.Z. A New Implementation of ontostep: Flexible Generation of Ontology and Knowledge Graphs of express-Driven Data. ASME. J. Comput. Inf. Sci. Eng., 22(2), 2022: 024502 (<http://doi.org/10.1115/1.4053079>)

# Открытый пример свежей реализации

- **NIST STEP to OWL Translator (STP2OWL)** (<http://github.com/usnistgov/STP2OWL>)
  - “STP2OWL translates STEP schemas (EXPRESS) and instance files (P21) into Web Ontology Language (OWL) files in a faster and more flexible way, thus furthering the adoption of the full capabilities of ISO 10303”

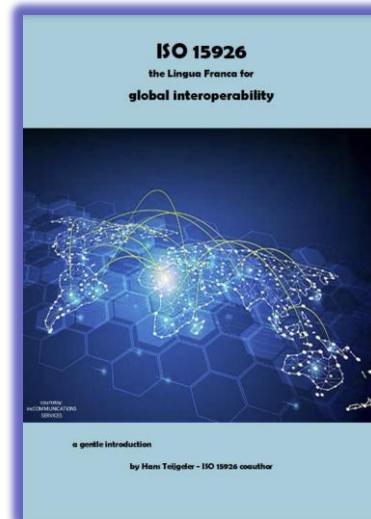


# Пример: онтологии в медицине

- Основные метаонтологии:
  - Глобальный медицинский тезаурус **Unified Medical Language System (UMLS) Metathesaurus** (<http://www.nlm.nih.gov/research/umls/quickstart.html>)
  - Интегратор источников данных и метаонтологий **SNOMED Clinical Terms (SNOMED CT)** (<http://www.snomed.org>)
  - **Medical Subject Headings (MeSH)** (<http://www.nlm.nih.gov/mesh/meshhome.html>)
  - **Medical Dictionary for Regulatory Activities (MedDRA)** (<http://www.meddra.org>)
- Открытые предметные онтологии:
  - Международная классификация болезней и проблем, связанных со здоровьем **International Classification of Diseases (ICD)** (<http://www.who.int/classifications/icd/en/>)
  - **The Disease Ontology** (<http://disease-ontology.org>)
  - Нормализованные названия лекарственных препаратов **RxNorm** (<http://www.nlm.nih.gov/research/umls/rxnorm/index.html>)
    - RxNav is a browser for several drug information sources, including RxNorm, RxTerms and NDF-RT (<http://mor.nlm.nih.gov/RxNav/>)
  - **LOINC is a freely available international standard for tests, measurements, and observations** (<http://loinc.org>)
    - Федеральный справочник лабораторных исследований (**ФСЛИ**). Справочник лабораторных тестов (<http://nsi.rosminzdrav.ru/#!/refbook/1.2.643.5.1.13.13.11.1080>)
    - Справочник лабораторных тестов и услуг (**ЛАТЕУС**) (<http://spbmiac.ru/ehlektronnoe-zdravookhranenie/informacionno-spravochnye-sistemy/spravochnik-lateus/>)
  - **Hymenoptera Anatomy Ontology Portal** (<http://api.hymao.org/projects/32/public/site/hymglossary/home/about>)
  - **HL7 Reference Information Model** (<http://www.hl7.org/implement/standards/rim.cfm>)
  - ...

# Пример: от ISO 15926 к OSDU (1)

- Серия международных стандартов для:  
“Integration of life-cycle data for process plants, including oil and gas production facilities”
  - Wikipedia – ISO 15926 ([http://en.wikipedia.org/wiki/ISO\\_15926](http://en.wikipedia.org/wiki/ISO_15926))
- 15926.org (<http://15926.org>)
  - *The home of an industrial data modelling team implementing the ISO 15926 interoperability standard*
  - *ISO 15926 – the Lingua Franca of global interoperability*  
(<http://15926.org/topics/ISO15926-consolidating-paper-20211219.pdf>)
- Пример TLO:
  - Batres R., West M., Leal D., Price D., Naka Y. An upper ontology based on ISO 15926, Computer Aided Chemical Engineering, 20, 2005, pp.1543-1548



# Пример работы 15926 browser

## 15926browser

RDL (Ref data Lib) ▾

data management

Deprecated objects

Provenance attributes

Search literal term

### DATA MANAGEMENT

**id**

<http://data.15926.org/rdl/RDS10804384>



**rdfs:label**

DATA MANAGEMENT

**rdf:type**

[ClassOfActivity](#)

**rdf:type**

[ACTIVITY OR EVENT](#)

**rdfs:subClassOf**

[MANAGING](#)

**skos:definition**

<DATA MANAGEMENT> is <MANAGING> of acquisition, validation, storage, protection, and processing, by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users

**prov:idISO**

12016

**prov:idPCA**

RDS10804384

**prov:statusPCA**

Recorded

**dcterms:creator**

u82237

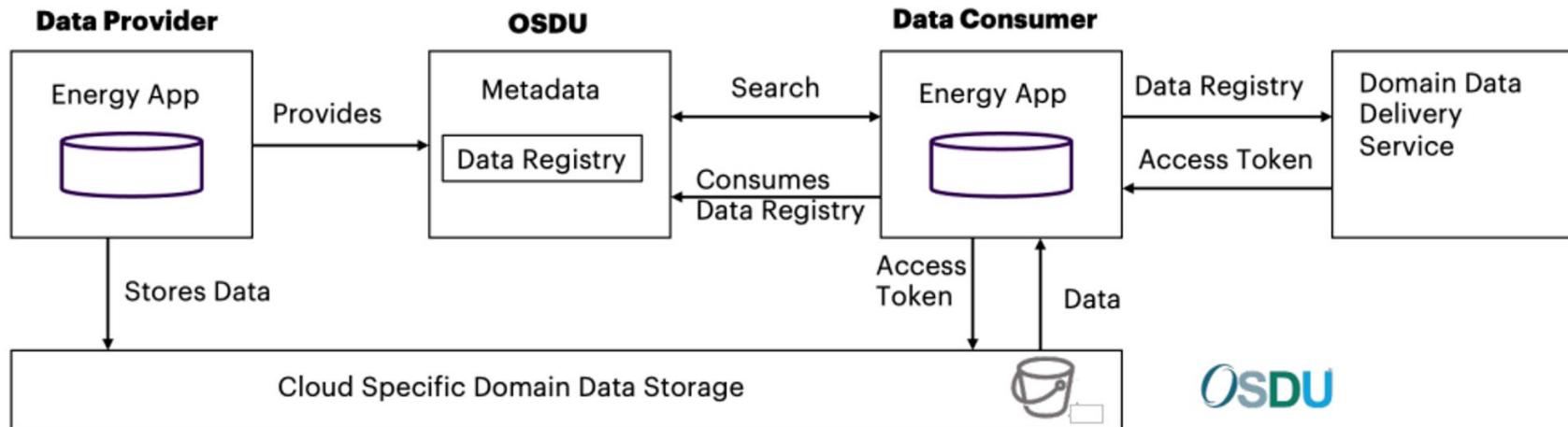
**meta:valEffectiveDate**

2005-02-01Z

# Пример: от ISO 15926 к OSDU (2)



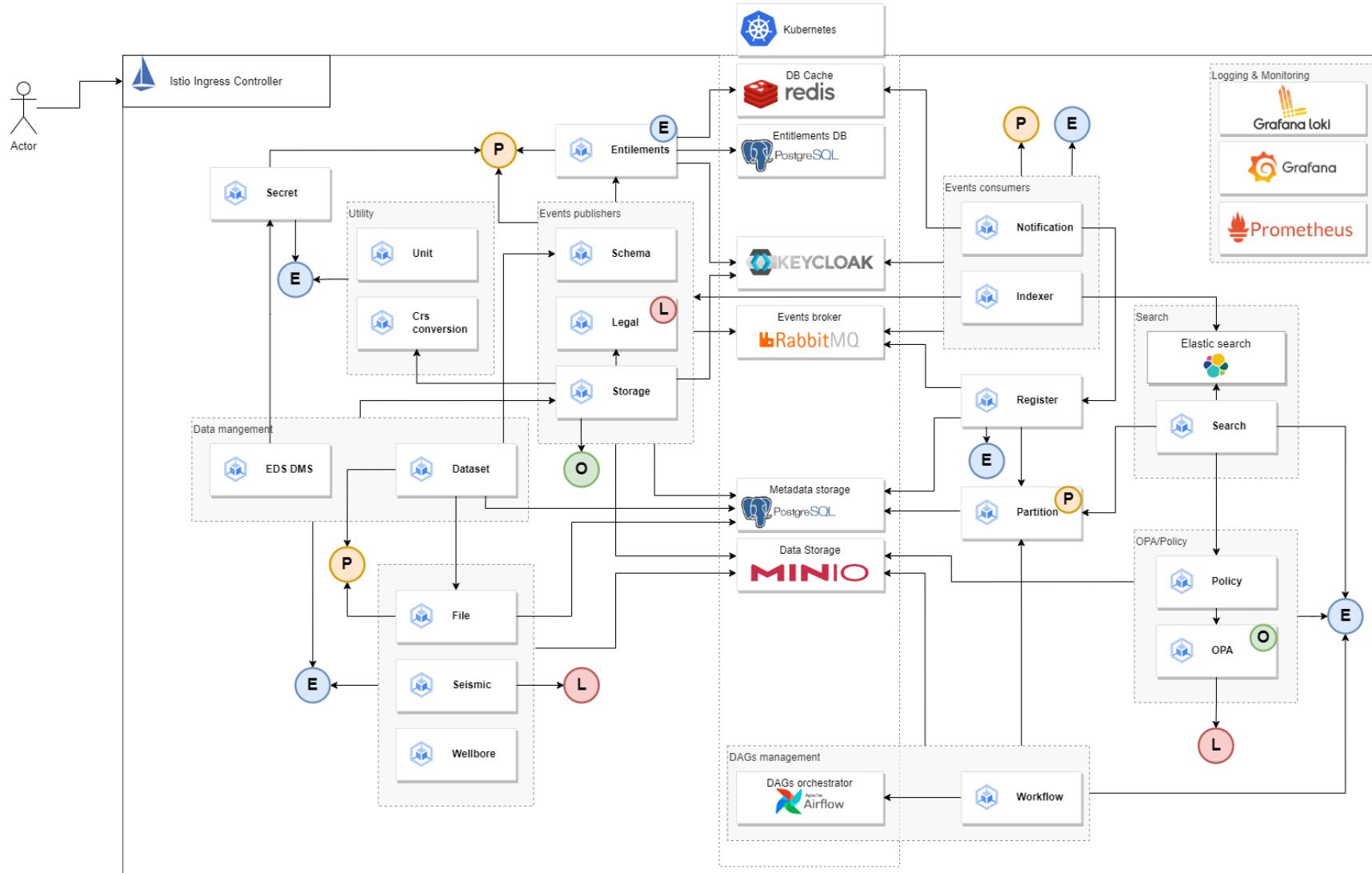
- ***Open Subsurface Data Universe*** (<http://osduforum.org>)
  - *OSDU Ontology* (<http://github.com/Accenture/OSDU-Ontology>)



OSDU

- Открытый *OSDU Deployment Framework*  
(<http://community.opengroup.org/osdu/platform/deployment-and-operations/infra-gcp-provisioning/>)

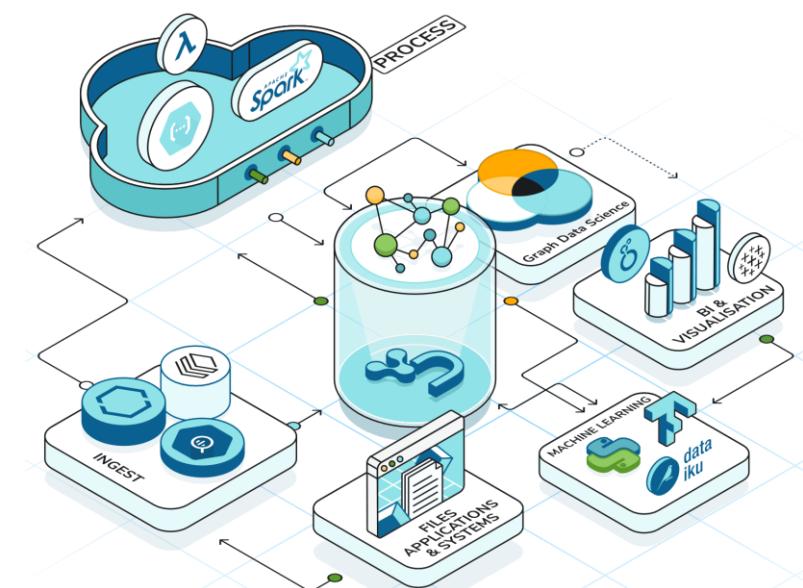
# Инфраструктурные компоненты OSDU Baremetal



# Главная открытая СУБЗ



- **Neo4j Graph Database** (<http://neo4j.com/product/neo4j-graph-database/>)
  - *Enterprise Edition for scalability, availability, and fine-grained access controls!*
- + **Neo4j Graph Data Science** (<http://neo4j.com/product/graph-data-science/>)
  - *Analytics and machine learning (ML) solution that analyzes relationships in data to improve predictions and discover insights*



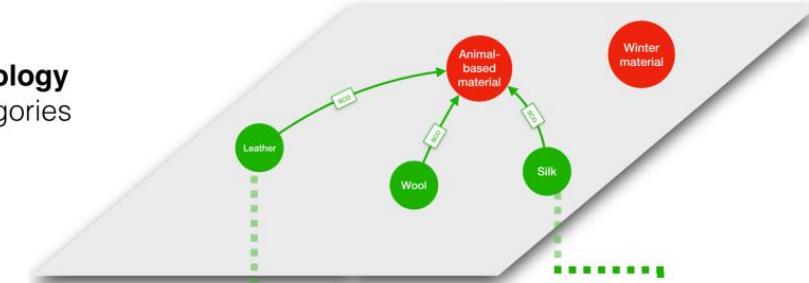
# *Neo4j Neosemantics*



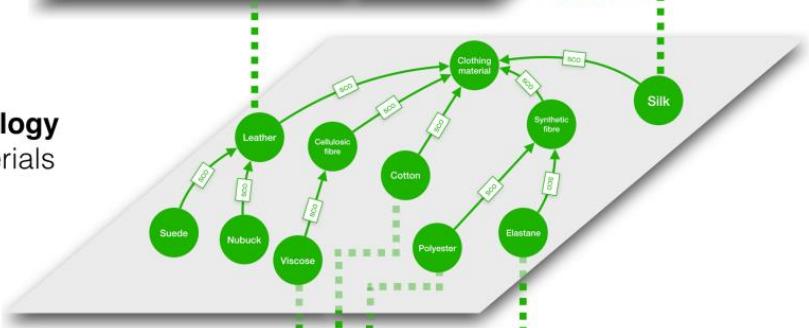
- ***neosemantics (n10s): Neo4j RDF & Semantics toolkit***  
[\(http://neo4j.com/labs/neosemantics/\)](http://neo4j.com/labs/neosemantics/)
  - *Plugin that enables the use of RDF and its associated vocabularies like (OWL, RDFS, SKOS and others) in Neo4j*
- Barrasa J. QuickGraph#9 The fashion Knowledge Graph.  
Inferencing with Ontologies in Neo4j. 2019  
[\(http://jbarrasa.com/2019/11/25/quickgraph9-the-fashion-knowledge-graph-inferencing-with-ontologies-in-neo4j/\)](http://jbarrasa.com/2019/11/25/quickgraph9-the-fashion-knowledge-graph-inferencing-with-ontologies-in-neo4j/)

# Example from QuickGraph#9

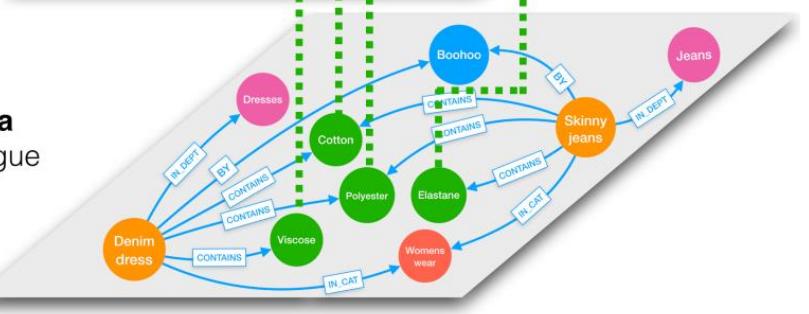
**Custom Ontology**  
Custom Categories



**Domain Ontology**  
Clothing Materials



**Instance Data**  
Product Catalogue



Graph interface showing navigation and search results:

- Graph tab selected.
- Search results: \* (8), Item(1), Class(2), Resource(2), Category(1), Department(1), Brand(1).
- Search results: \* (5), CONTAINS(2), IN\_CAT(1), IN\_DEPT(1), BY(1).

Graph visualization:

```

graph TD
    Menswear((Menswear)) -- IN_DEPT --> ONeillLogoPrintTee((O'Neill Logo Print Tee))
    ONeillLogoPrintTee -- CONTAINS --> Cotton((Cotton))
    ONeillLogoPrintTee -- CONTAINS --> Viscose((Viscose))
    ONeillLogoPrintTee -- IN_CAT --> TShirts((T-Shirts))
    ONeillLogoPrintTee -- BY --> ONeill((O'Neill))
  
```

Table view (Search results):

"White Stuff Blue Issy Wedge Shoes"	"White Stuff"	"Main - 100% Suede, sole - 100% Other fibres."
"Brakeburn Ditsy Sandals"	"Brakeburn"	"Upper: 100% Polyvinylchloride, Lining: 100% Felt, Insole: 100% Suede, Outsole: 100% EVA."
"Brakeburn Shells Crossover Sandals"	"Brakeburn"	"Upper: 100% PVC, Lining: 100% Felt, Insole: 100% Suede, Outsole: EVA."
"Brakeburn Lemon Sandals"	"Brakeburn"	"Main: 100% Polyvinylchloride, Lining: 100% Other Fibres, Sole: 100% Suede, Outsole: EVA."
"Call It Spring Ladies Heeled Suede Sandals"	"Call It Spring"	"Upper: 100% Nubuck, Sole: 100% Thermoplastic Rubber."
"Gabor Black Shiraz Foot Fit Nubuck Mid Leg Boots"	"Gabor"	"Main - 100% Nubuck, Lining - 100% Other fibres, Sole - 100% Rubber."
"Skagen® Signature Watch"	"Skagen"	"100% Leather"
"Skagen® Signature Strap Watch"	"Skagen"	"100% Leather"
"Skechers® Flex Advantage 3.0"	"Skechers"	"Upper: Leather, Lining and Sock - Leather, Sole - Other Materials."
"Skechers® Go Walk 5 Slip-On Shoes"	"Skechers"	"Upper: Leather, Lining and Sock - Textile, Sole - Other Materials."

Diagram view (Bottom right):

```

graph TD
    Leather((Leather)) -- IS-A --> Nubuck((Nubuck))
    Leather -- IS-A --> Suede((Suede))
  
```

# А управление метаданными?



- **OpenMetadata** (<http://open-metadata.org>)
  - A Single Place to **Discover**, Collaborate and get your Data Right
  - Business Glossary enhancements (v.1.4):
    - **Custom Properties**: Custom Properties enable teams to enrich data asset metadata by adding additional attributes. Previously, we supported primitive types such as text, integer, and string. In this release, we've expanded the available types to include Enums, **Entity References**, Date, Time, and more
    - These new types allow users to **link other assets** within OpenMetadata or create Enums, providing greater flexibility and depth in metadata management



# Корпоративные KMS – *PoolParty*



- **Semantic Web Company *PoolParty Semantic Suite***

(<http://semantic-web.com/poolparty-semantic-suite/>)

- **How *PoolParty* and ISO 25964 fit together**

(<http://semantic-web.com/2016/03/04/how-poolparty-and-iso-25964-fit-together/>)

SEMANTIC WEB COMPANY

CONTACT US

PRODUCT SERVICES INNOVATION CAREER CUSTOMERS SEMANTIC INSIGHTS COMPANY

SEARCH

The Most Complete Semantic Middleware

PoolParty Semantic Suite is world-class technology for your data.

WHAT WE CAN BUILD TOGETHER

**Grow your business with better systems.**

Altogether, our semantic capabilities allow you to transform outdated, siloed systems into intelligent applications that use your data more effectively. Benefit from a more productive workforce and a happier customer base.

**TAXONOMIES**  
Build the foundations of your knowledge model with a structure that your machine can understand.  
[read more →](#)

**TEXT MINING**  
Use natural language processing to extract the most valuable pieces of your data and content.  
[read more →](#)

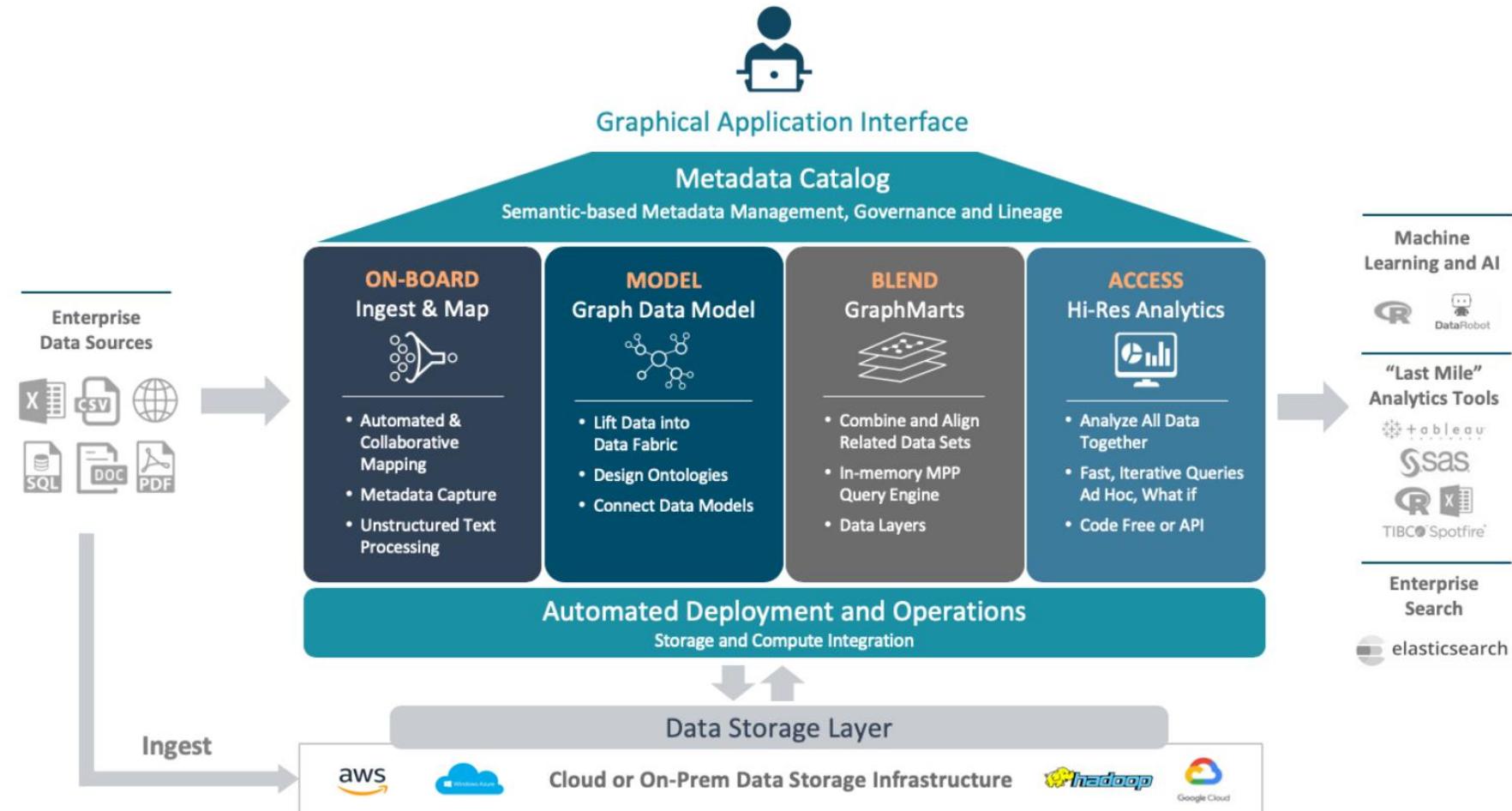
**KNOWLEDGE GRAPHS**  
Connect business assets and data silos in a context-rich ecosystem. Generate holistic views.  
[read more →](#)

**ENTERPRISE SEARCH**  
Access information more easily and get tailored content recommendations.  
[read more →](#)

# Корпоративные KMS – Anzo



- *Cambridge Semantics Anzo* (<http://www.cambridgeseantics.com/product/>)



# Корпоративные KMS – *Stardog*

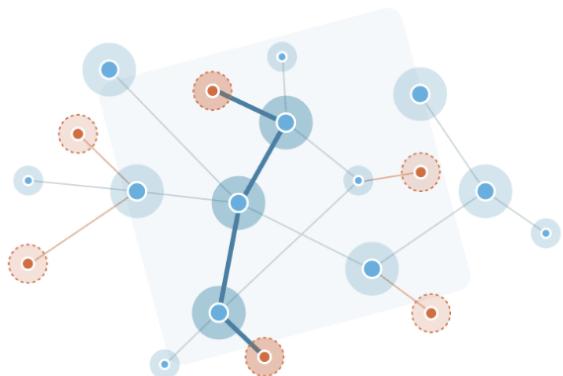


- *Stardog – Turn your Data into Knowledge... Fast* (<http://www.stardog.com>)

## Knowledge Graphs = Graph + AI + Virtualization

### Based on graph

Graph's flexibility alone is not enough to turn data into knowledge. **Graph databases** — a common use of graph — can accept new data more easily than relational databases, but functionality is limited by its single schema. But a graph database imposes one point of view of the world and requires that business logic is coded into the application directly, whereas the low-code Knowledge Graph stores logic centrally.



Turn data into knowledge with Stardog's best-in-class Inference Engine

Stardog's Inference Engine associates related information stored in disparate sources, and then uses this rich web of relationships to discover new relationships within your data. By inferring new connections between concepts in the Knowledge Graph, the resulting network of information becomes increasingly more valuable. Furthermore, this represents the breadth of organizational knowledge in a machine-understandable format. With knowledge digitized, Knowledge Graphs support better decision-making and faster time to insight.

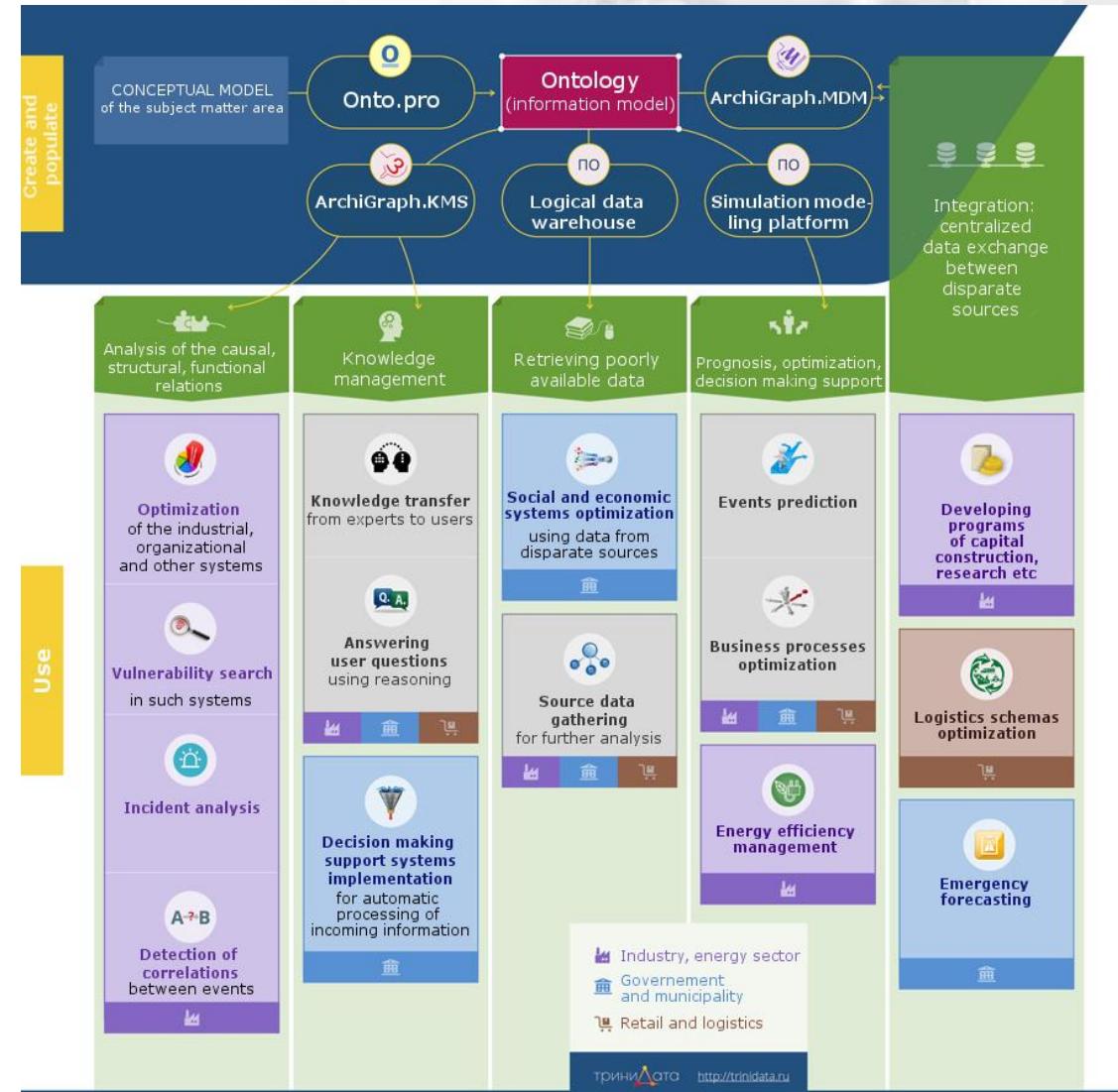
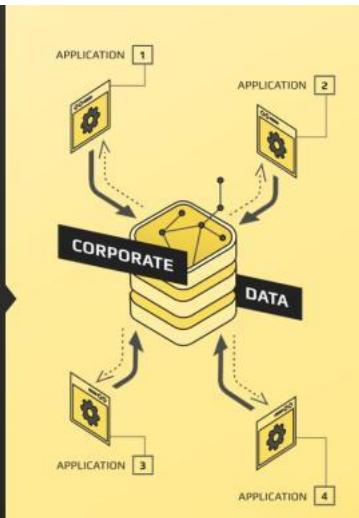
Inference creates new relationships by interpreting your source data against your data model. By expressing all the implied relationships and connections between your data sources, you create a richer, more accurate view of your data. This includes the ability to represent multiple definitions for the same data, empowering collaboration between stakeholders and expressing situational or theoretical truths.

- 01 Connect and query data of any structure
- 02 Virtualize or materialize
- 03 Inference Engine for explainable AI**
- 04 Built-in machine learning
- 05 Navigate your data with Pathfinder
- 06 Data quality management
- 07 Based on open standards
- 08 High-performance graph database

# Импортозам – *Trinidata* (1)

- ТриниДата (<http://trinidata.ru>)
  - Горшков С. Введение в онтологическое моделирование.  
Тринидата, 2016 (<http://trinidata.ru/files/SemanticIntro.pdf>)
  - Горшков С.В. Онтологическое моделирование предприятий:  
методы и технологии. Изд-во Урал. ун-та, 2019. 236 с  
(<https://trinidata.ru/files/EnterpriseModeling.pdf>)
  - ООО «Тринидата». Корпоративные автоматизированные системы на основе  
онтологических моделей: книга рецептов. 2020 (<https://trinidata.ru/files/CookBook.pdf>)

# Импортозам - *Trinidata* (2)



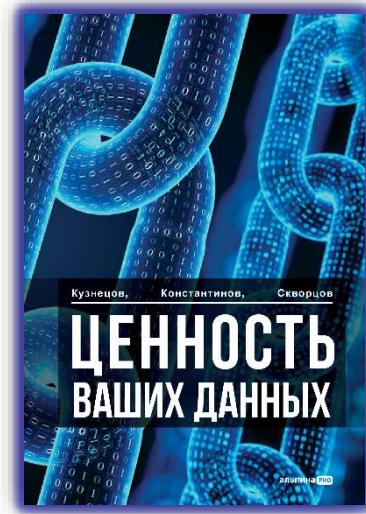
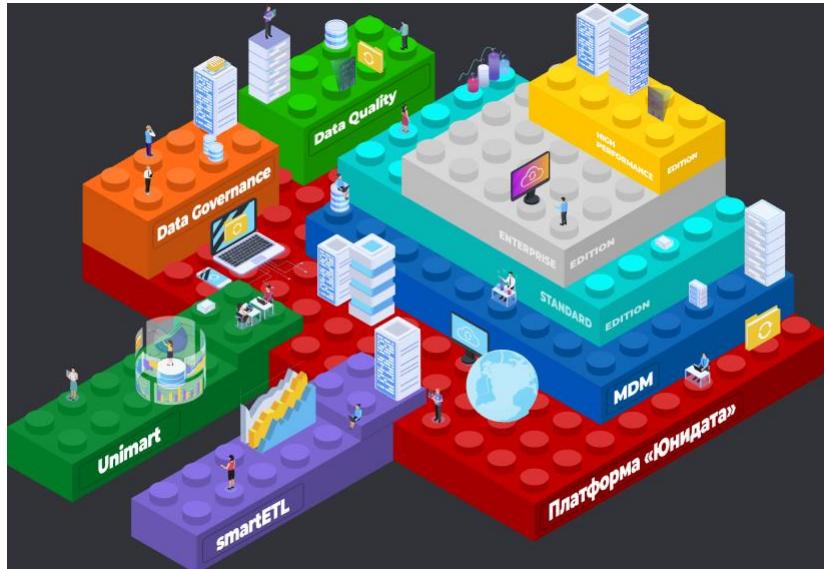
# Импортозам – *Unidata*



- Юнидата (<http://unidata-platform.ru>)

- *Unidata Community* – объединение экспертов в области управления данными (<https://unidata-community.ru>)

- Кузнецов С., Константинов А., Скворцов Н.  
Ценность ваших данных. Альпина PRO, 2022.



# Выводы

- Всё называется «не так» и всё время появляются «**птичьи языки**»
  - В теории графов, онтологии и анализе формальных понятий уже почти всё есть 😊
- Корпоративные системы в основе в той или иной мере «**онтологизированы**»
  - Соответственно, все эволюционно развиваются решения
- Серьёзные системы управления знаниями **очень дороги**
  - Поэтому живут несколько «**миров**», включая академический
- Гибридизация моделей ИИ и получаемые преимущества хорошо **встряхнули рынок** и революция в корпоративном секторе на подходе
- Чисто технически:  $XML \rightarrow JSON$ ,  $OWL \rightarrow RDF + JSON-LD + SHACL, \dots, SQL + GQL$
- Гибридизация моделей ИИ и интеллектуализация ввода/выводы (**БЯМ/LLM**)

# Конец первой части?!

**Вопросы? Замечания? Предложения?**

