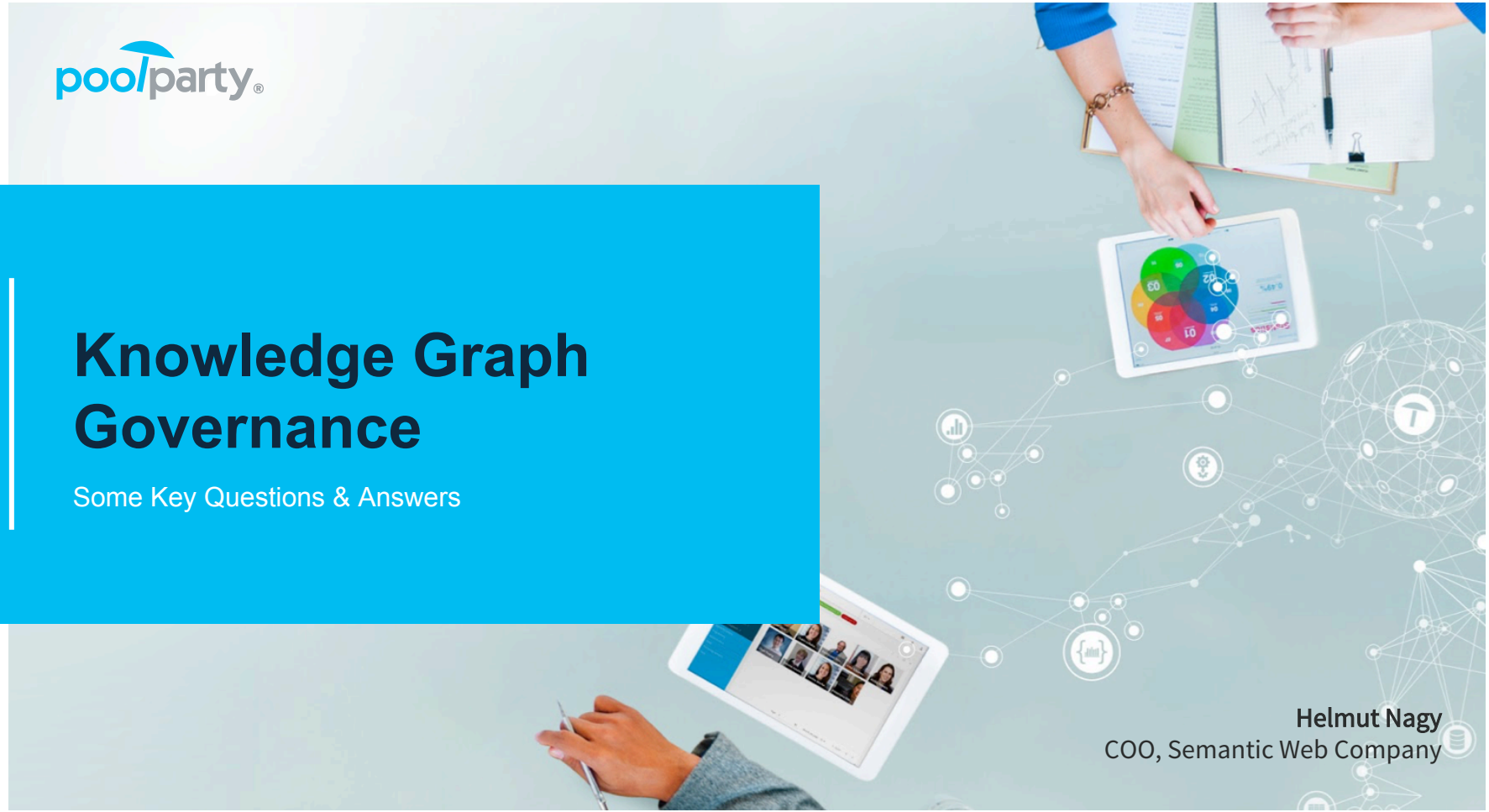
Knowledge Graph Governance

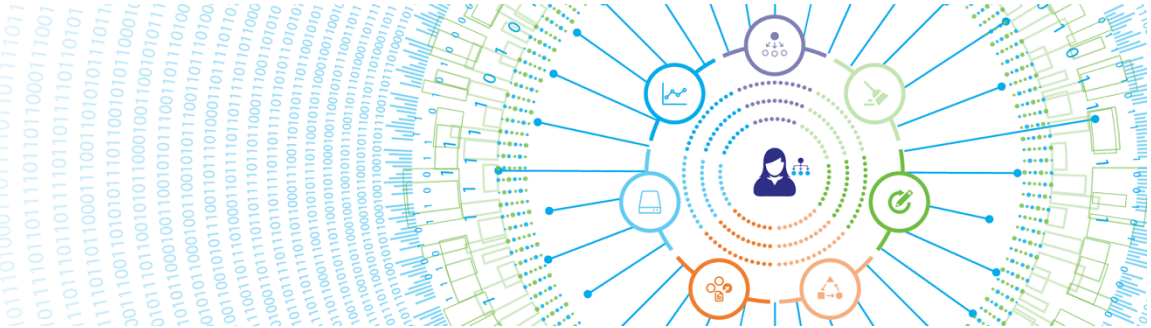
Some Key Questions & Answers



Helmut Nagy
COO, Semantic Web Company

What is a

KNOWLEDGE GRAPH



“The Knowledge Graph is a knowledge base used by Google and its services to enhance its search engine's results with information gathered from a variety of sources.” (Wikipedia)

"For knowledge graphs in information science, see https://en.wikipedia.org/wiki/Ontology_(information_science)"

“In computer science and information science, an ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse.” (Wikipedia)

"... Translating research papers within every field is a problem made easier when experts from different countries maintain a controlled vocabulary of jargon between each of their languages"

“A Knowledge Graph is a model of a knowledge domain created by subject-matter experts with the help of intelligent machine learning algorithms. It provides a structure and common interface for all of your data and enables the creation of smart multilateral relations throughout your databases.” (SWC)

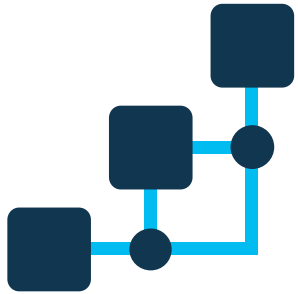
“A knowledge graph is unified information across an organization, enriched with contextual and semantic relevance across the silos. It combines capabilities of graph data stores with a knowledge toolkit for data unification and provides a holistic view of the organization’s data through relationships.” (Gartner)

“A knowledge graph

1. *mainly describes real world entities and their interrelations, organized in a graph.*
2. *defines possible classes and relations of entities in a schema.*
3. *allows for potentially interrelating arbitrary entities with each other.*
4. *covers various topical domain*

The first two criteria clearly define the focus of a knowledge graph to be the actual instances (A-box in description logic terminology), with the schema (T- box) playing only a minor role.” (Paulheim)

Building Blocks of a Knowledge Graph

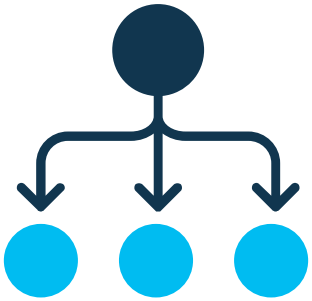


“A Knowledge Graph is a model of a knowledge domain”

Model means: "A representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities"

Building Block One: Conceptual/Domain model (Ontology)

Building Blocks of a Knowledge Graph

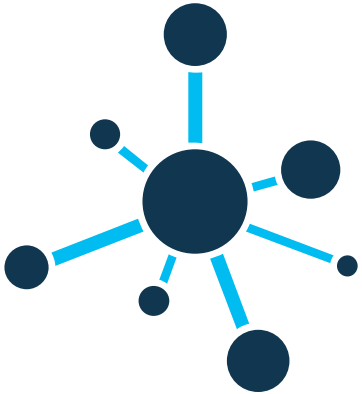


“A Knowledge Graph must eliminate ambiguity”

Eliminate ambiguity introduced by language. This also can cover multilinguality.

Building Block Two: Controlled Metadata (Vocabulary/Taxonomy)

Building Blocks of a Knowledge Graph

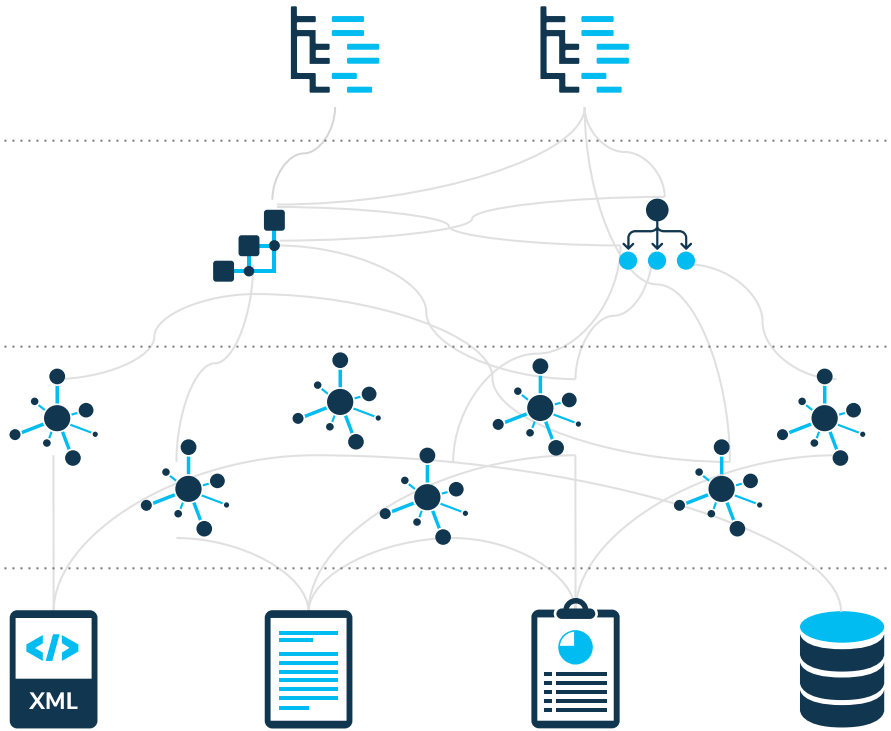


"A Knowledge Graph is unified information across an organization"

Unified information across an organization, enriched with contextual and semantic relevance across the silos.

Building Block Three: (Virtual) Data Layer

Form a Knowledge Graph



Augmented AI application

Conceptual model and Controlled Metadata

(Virtual) Data Layer

Content & Data Layer

Knowledge Graph is Explicit Knowledge

- Interpretation layer on top of existing data
- Connecting data silos
- Providing a unified access to structured and unstructured data
- Providing information in a machine readable way

Key to “Augmented Intelligence”

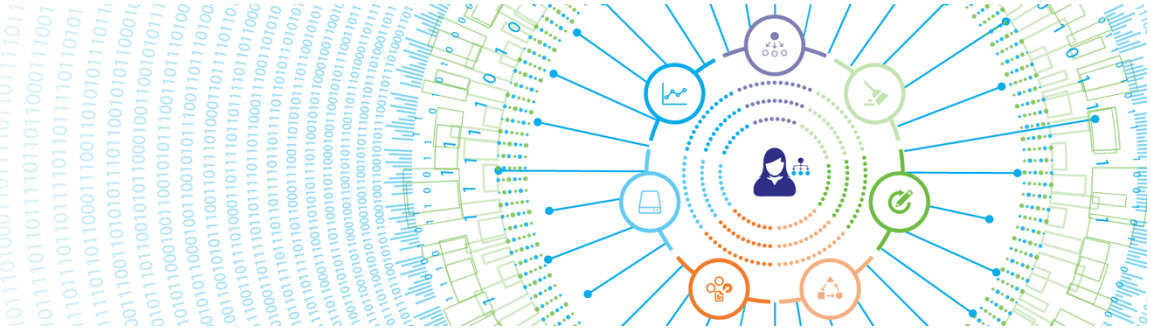
Augmented Intelligence (AI)

“The goal of AI should be to empower humans to be better, smarter and happier, not to create a “machine world” for its own sake. ...

People are the strongest component of AI. Smart people, not smart machines, develop the most sophisticated AI systems.” (Gartner)

Existing

KNOWLEDGE GRAPHS



The Google Knowledge Graph

Main use cases:

- Find the right thing
- Get the best summary
- Go deeper and broader

History:

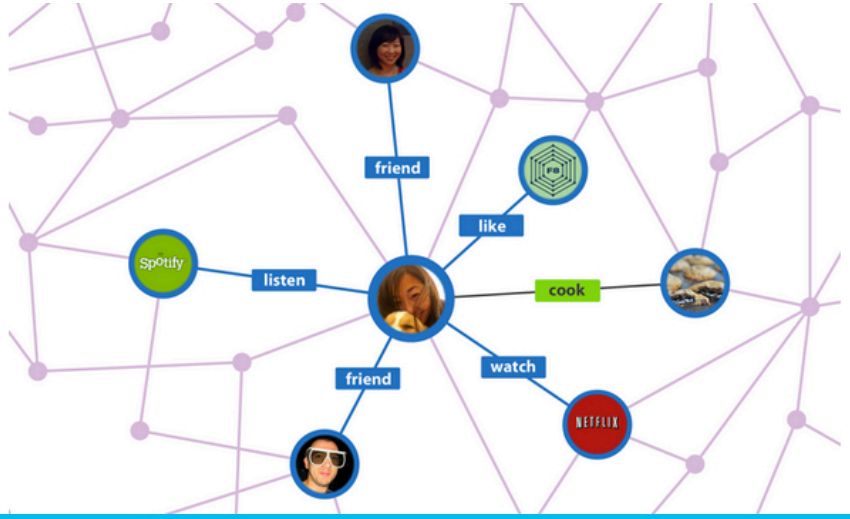
- Google acquired Freebase (2010)
- First Knowledge Graph in Google Search, only US (2012)
- Migration to Wikidata (2015)

World Knowledge Graph

The Facebook Graph Search aka Knowledge Graph

"Facebook Graph Search was a semantic search engine that was introduced by Facebook in March 2013. … The name refers to the social graph nature of Facebook, which maps the relationships among users."

Social Graph

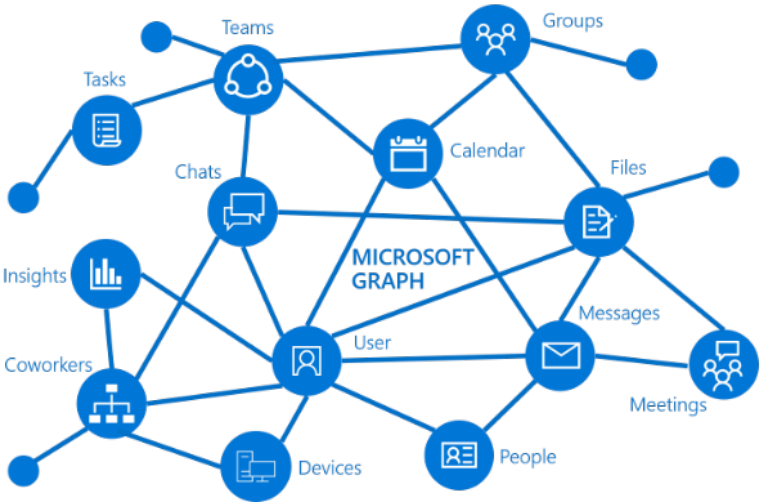


The Microsoft Graph API aka Knowledge Graph

"Microsoft Graph is the gateway to data and intelligence in Microsoft 365. It provides a unified programmability model that you can use to access the tremendous amount of data in Office 365, Windows 10, and Enterprise Mobility + Security. "

Office Graph

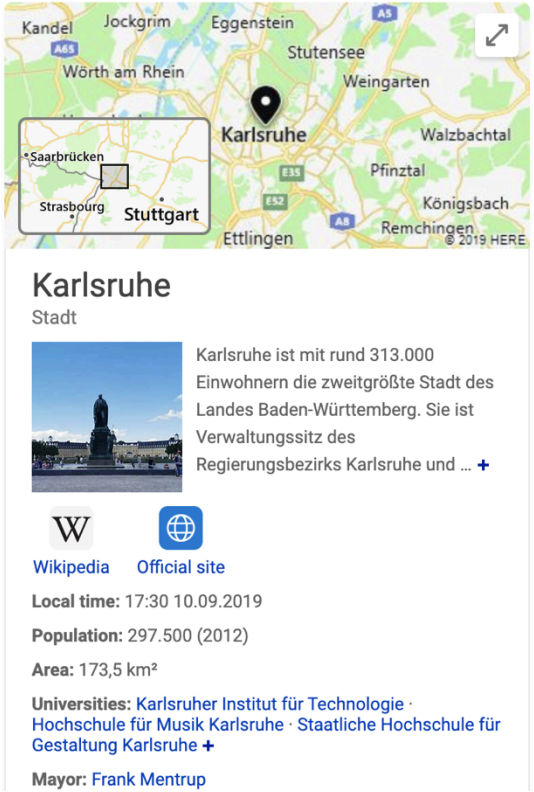
-> there is also MS Satori



Microsoft Satori

“Like Google's Knowledge Graph, Satori catalogs entities and the associated data and relationships among them. They both crawl the Web and utilize existing data sources, such as Freebase (owned by Google) and Wikipedia, to build their repositories of semantically rich encoded information that can help answer questions in milliseconds, rather than deliver a page of links.”

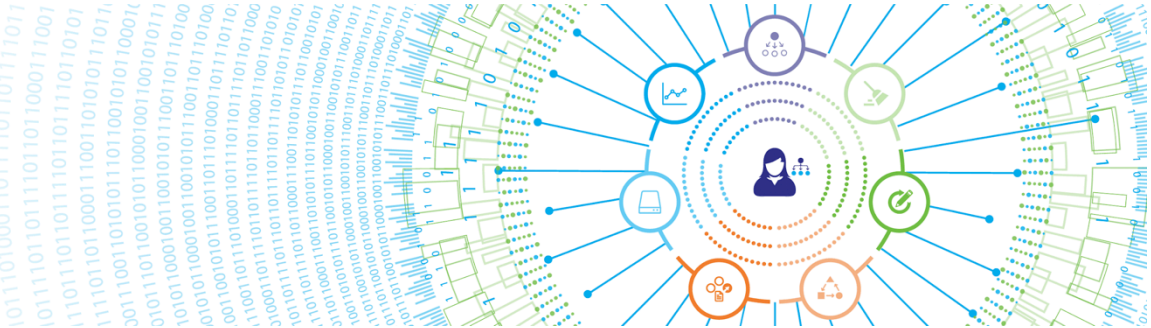
World Knowledge Graph



Does that help us in a specific enterprise/domain!

Knowledge Graph

Governance



A Knowledge Graph = Semantic Knowledge Graph

"Relational databases aren't designed to capture rich interrelationships across a large number of domain objects. …

Graph databases store connections as relationships. Relationships are first-class citizens in a graph data store. …

RDF triple stores are proficient at … building knowledge graphs and use cases where the domain requires relationships to be explored from all directions." (Gartner)

Basic Governance Principle

Follow basic Semantic Web Principles

- URIs are the key element of a knowledge graph

You need a URI strategy

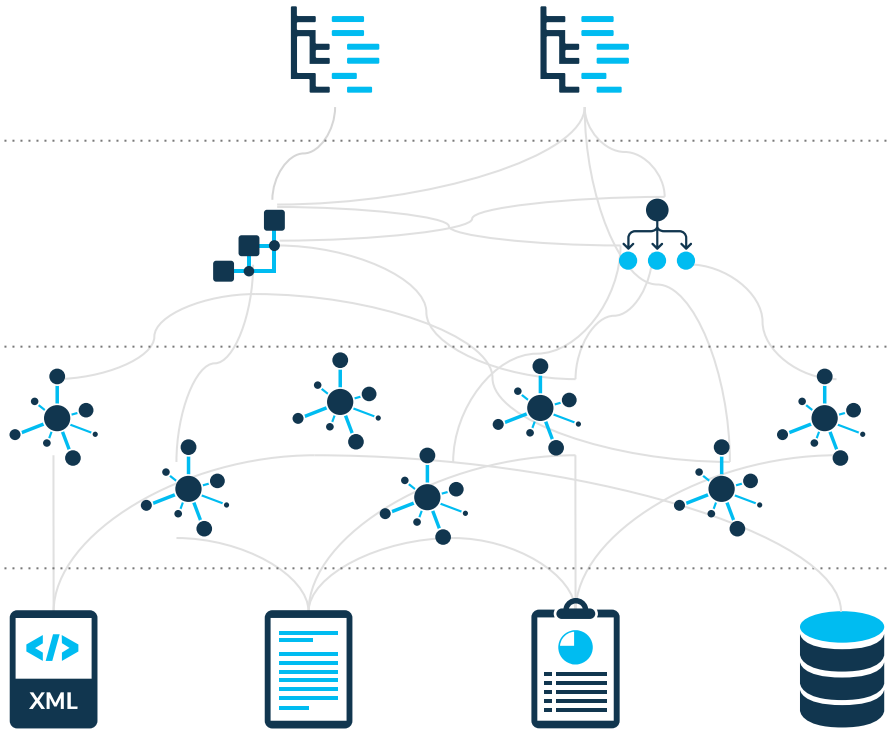
- Reuse of existing models and vocabularies is good practice

“Not invented” here is a good thing

- Knowledge Graphs are built to be extended and evolve

Start small and grow (based on use cases)

Knowledge Graph Governance



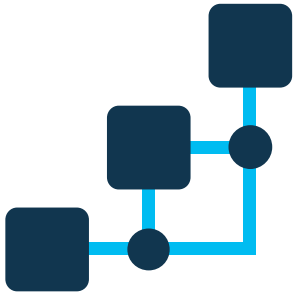
Augmented AI application

Conceptual model and Controlled Metadata

(Virtual) Data Layer

Content & Data Layer

Conceptual/Domain model



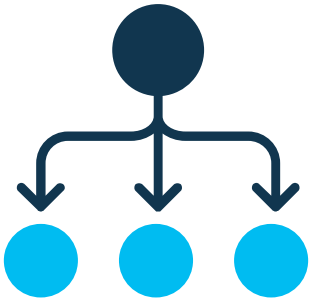
The model

- will change over time
- will evolve and be extended
- will be linked to other models

Changes to the model influences

- the controlled metadata layer
- the (virtual) data layer
- applications built on top
- creation of the (virtual) data layer

Controlled Metadata



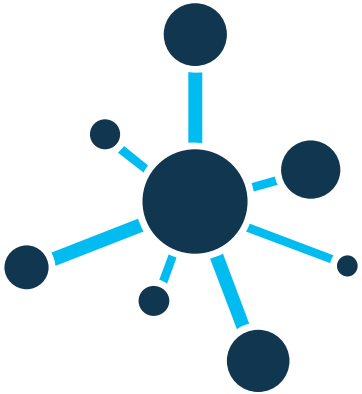
The metadata layer

- will change over time
- will evolve and be extended
- will be linked to other metadata

Changes to the metadata layer influence

- the (virtual) data layer
- applications built on top
- creation of the (virtual) data layer

(Virtual) Data Layer



The data layer

- will change over time
- will evolve and be extended
- new sources will be added

Changes to the data layer influence

- the model
- the controlled metadata layer
- applications built on top
- creation of the (virtual) data layer

Governance - Administrative metadata

- Versioning
- Workflows / Data life cycle
- History / Layers
- Provenance / Data owner ship

- On dataset level
- On data level

Connect

Helmut Nagy

COO, Semantic Web Company

- helmut.nagy@semantic-web.com
- https://www.linkedin.com/in/helmutnagy

Q & A