Francesco Beretta

CNRS UMR5190 LARHRA – Université de Lyon

Université de Neuchâtel

Linked Open Data for Research

Towards a Paradigm Shift in the Humanities and Social Sciences?

Linked Open Data in Dialogue

University of Bern, 15 November 2023

Linked Open Data and the Semantic Web

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data.

https://5stardata.info/en/

★ make your stuff available **on the Web** (whatever format) under an open licence

★★ make it available as **structured data** (e.g., Excel instead of image scan of a table)

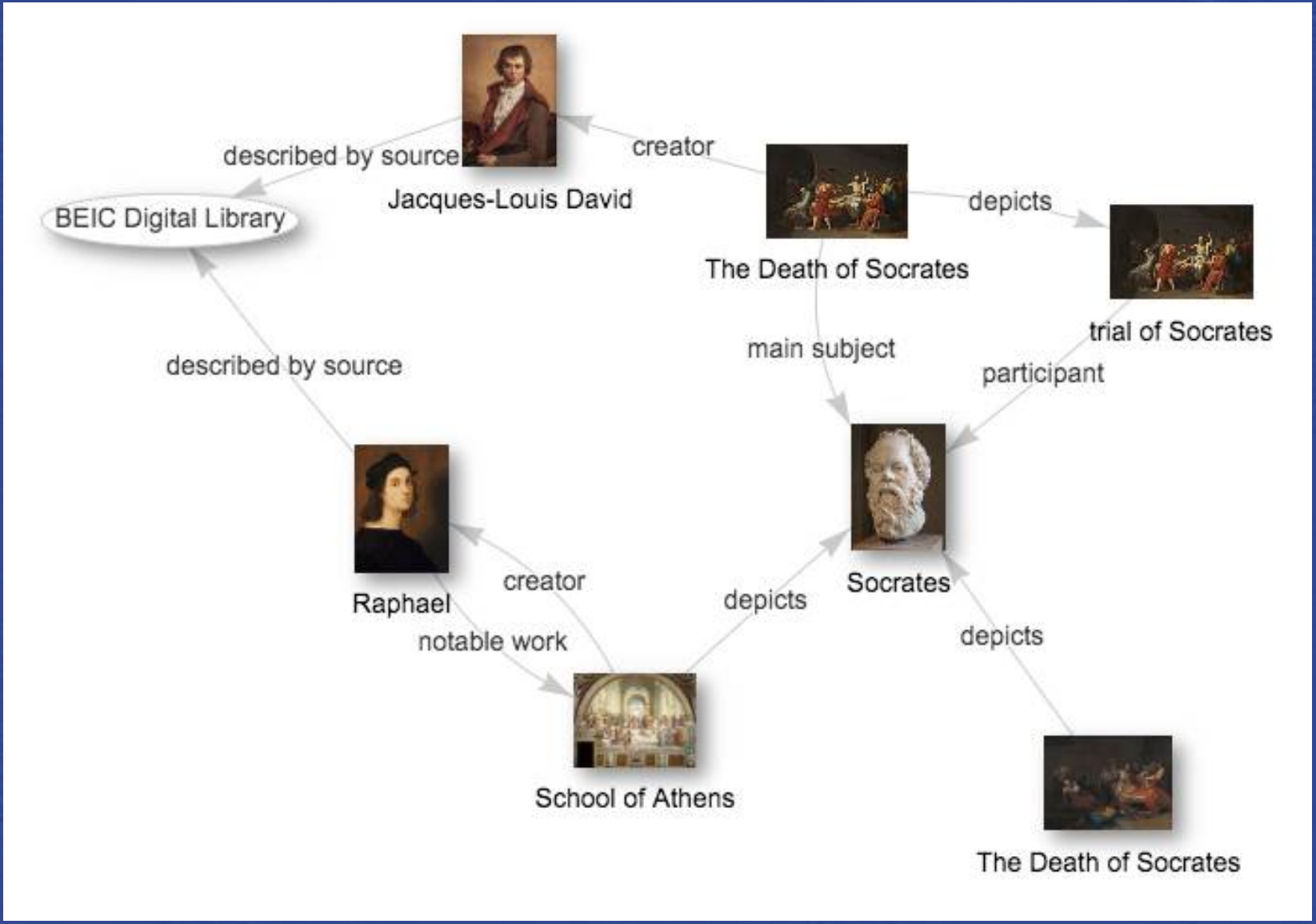
★★★ make it available in a non-proprietary **open format** (e.g., CSV instead of Excel)

★★★★ use **URIs to denote things**, so that people can point at your stuff

★★★★★ **link your data** to other data to provide context

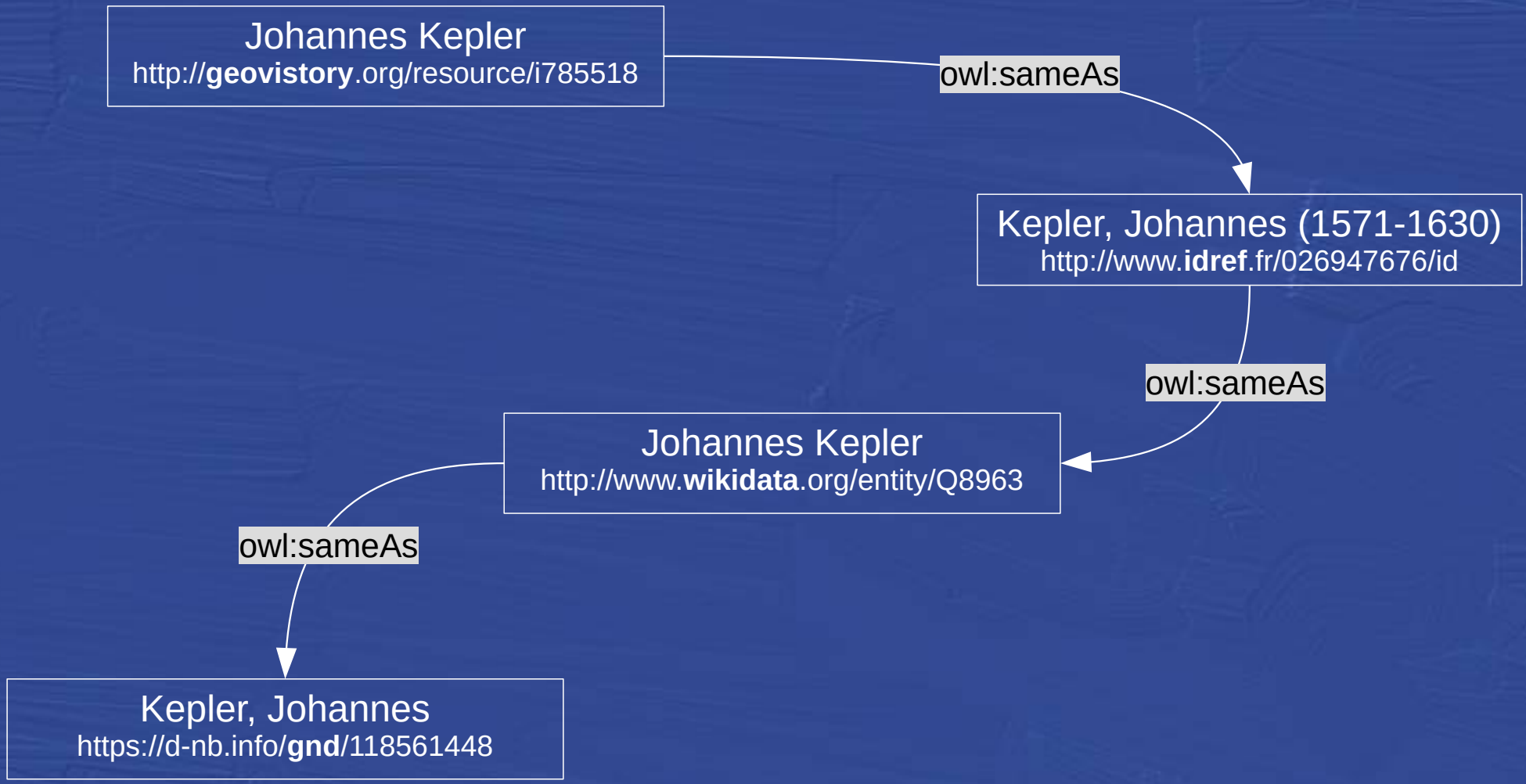
Quoted from : https://5stardata.info/en/

Wikidata : an information graph (*knowledge graph*) representing the objects in the world and their relationships



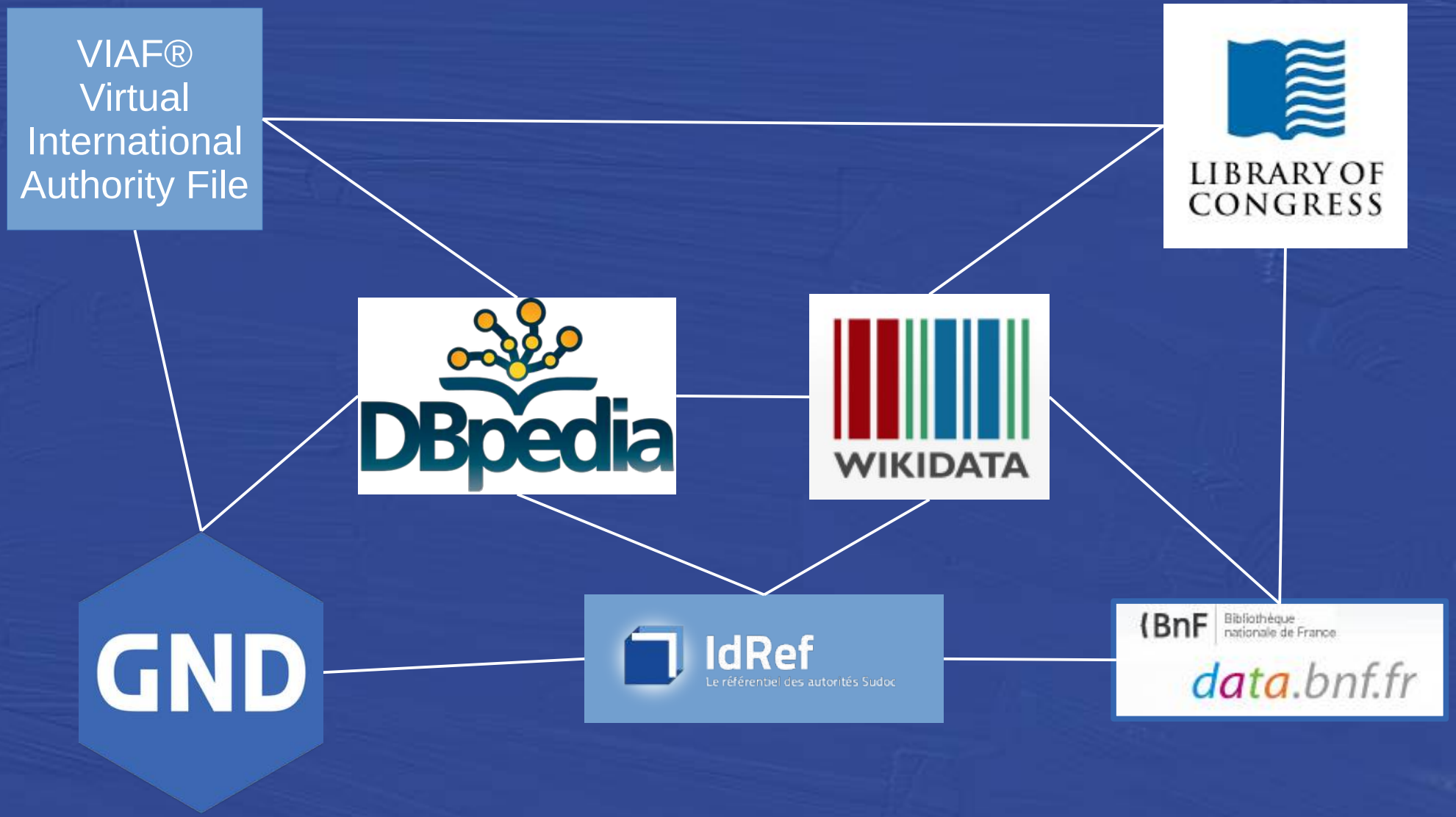
LOD as Links

« URIs to denote things, so that people can point at your stuff »

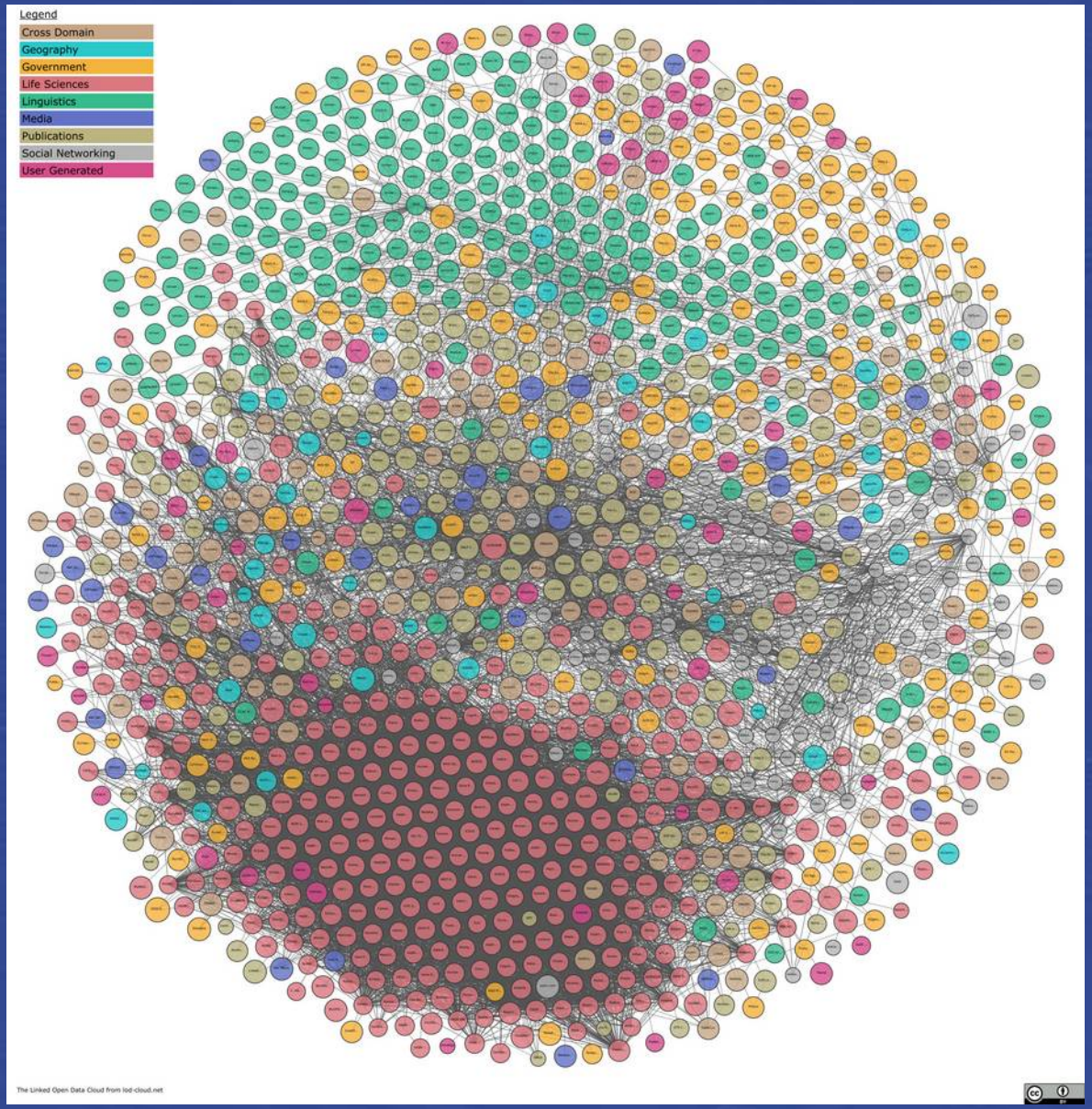


LOD as Links

« URIs to denote things, so that people can point at your stuff »

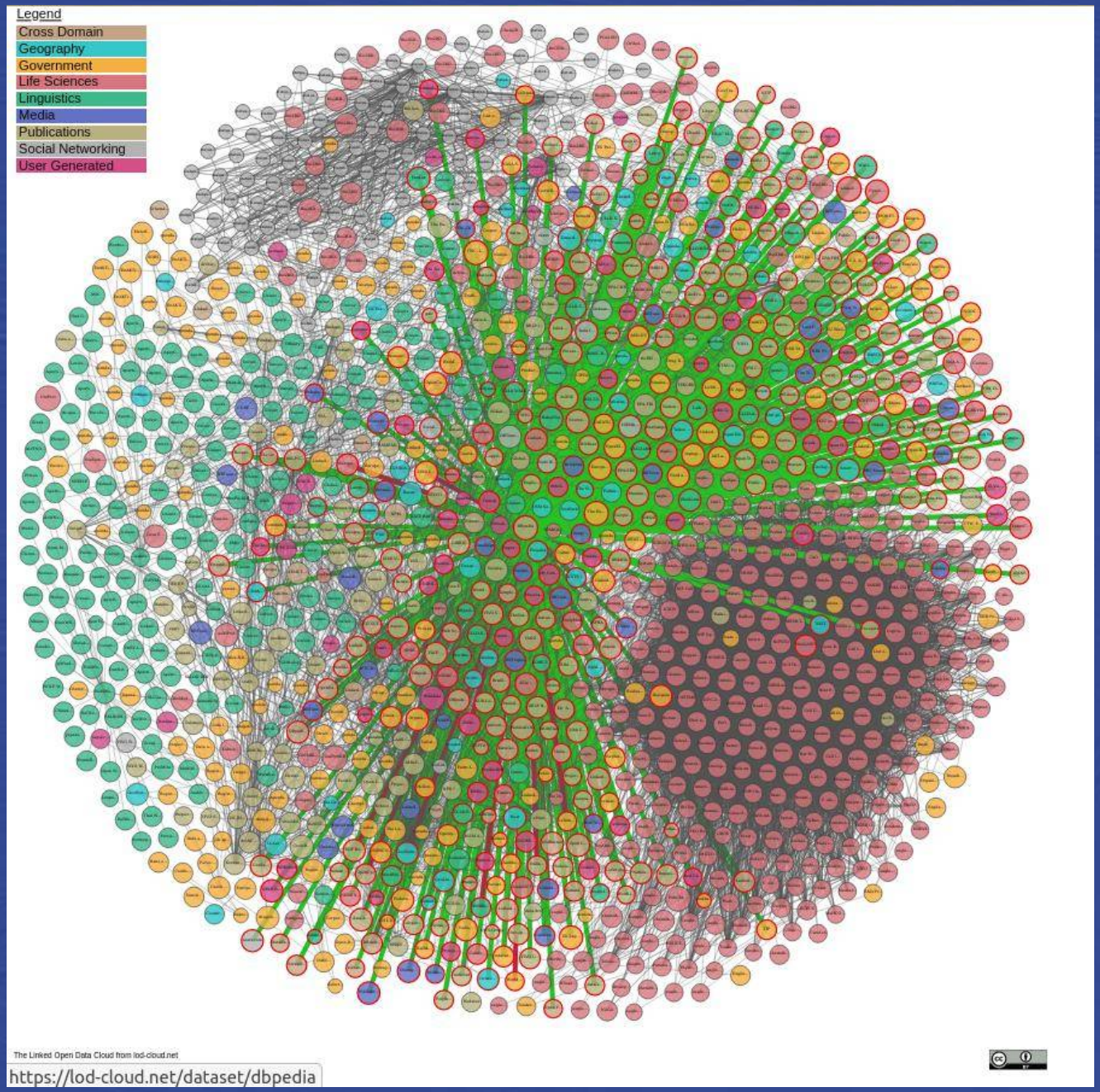


Web sémantique (Wikipédia)

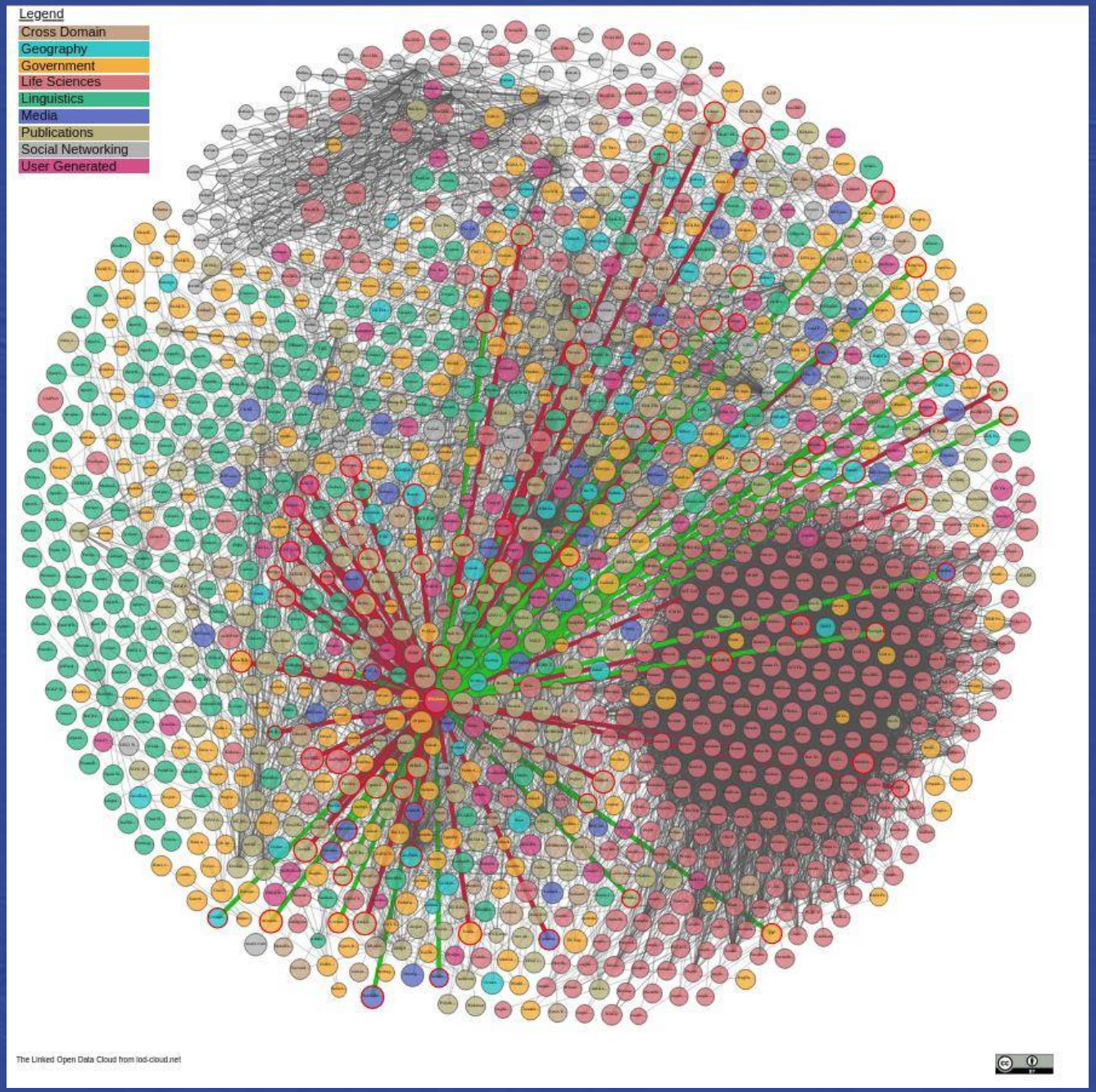


https://lod-cloud.net/

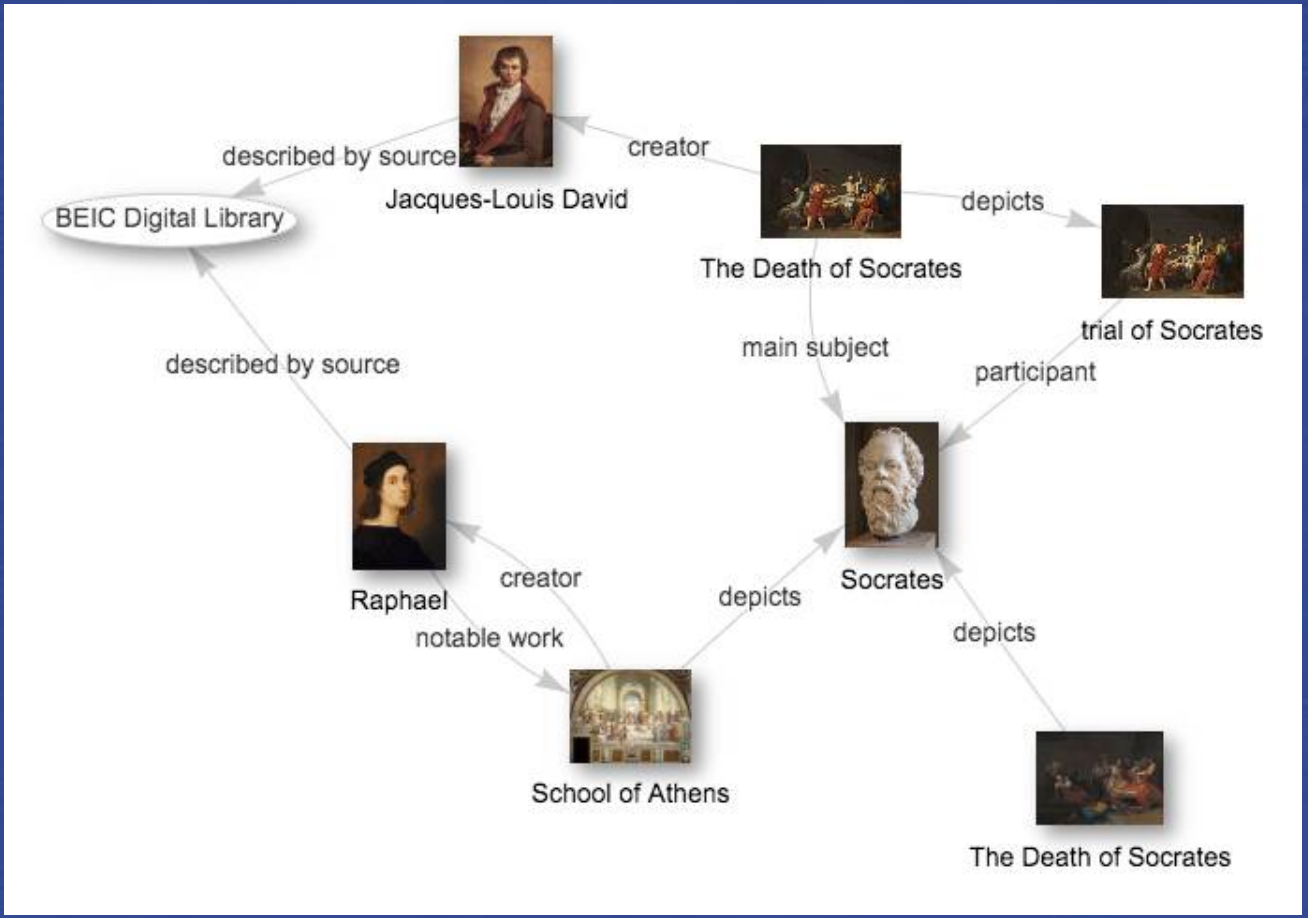
DBPedia


Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
The Linked Open Data Cloud from lod-cloud.net
https://lod-cloud.net/dataset/dbpedia

Wikidata


Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
The Linked Open Data Cloud from lod-cloud.net

LOD as a Semantic Graph

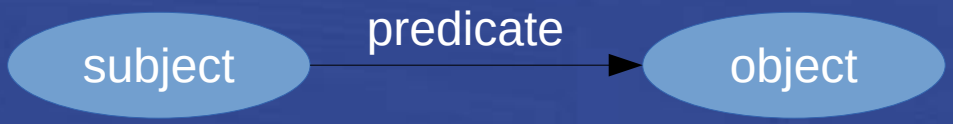


Linked Open Data and the Semantic Web

(https://www.w3.org/TR/rdf11-concepts/)

- « The Resource Description Framework (**RDF**) is a framework for **representing information in the Web**. »

- « A graph-based data model »

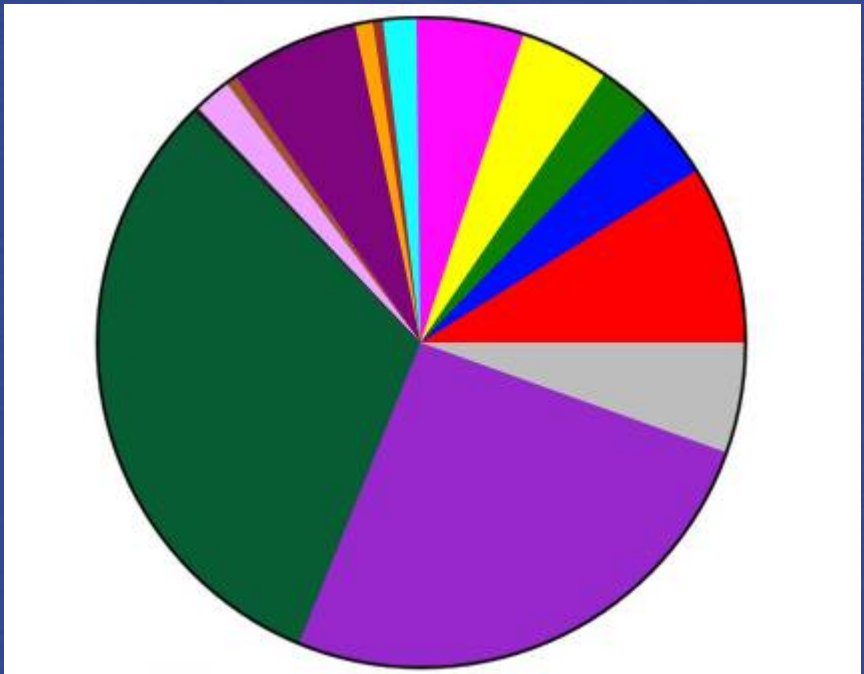


RDF
triple

Wikidata

14 November 2023 – 107’588’216 items

1.5 billions statements

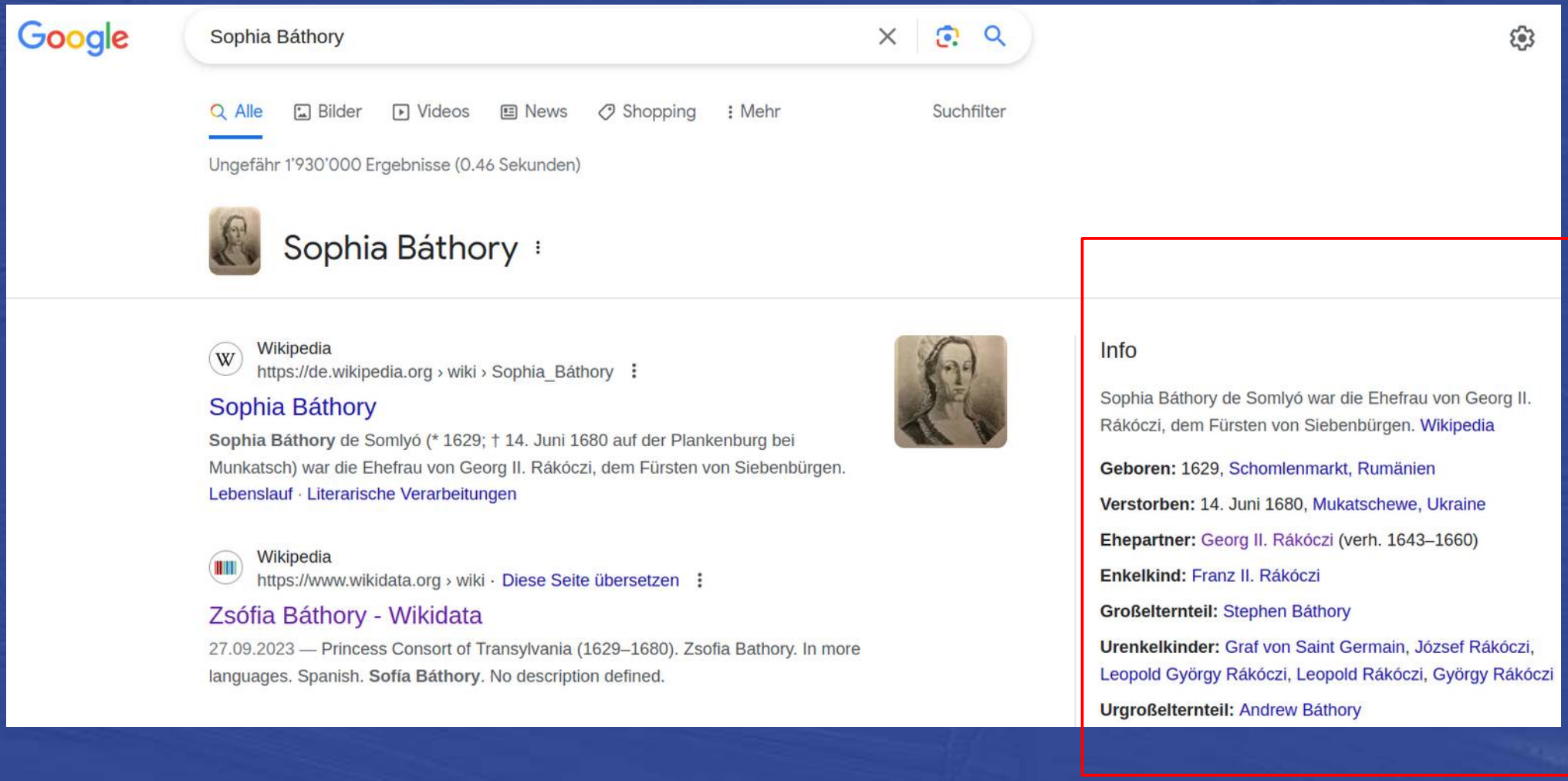


16 February 2020 : 71,611,020 items

human: 6,376,879 (8.9%)
taxon: 2,726,046 (3.8%)
administrative division: 1,943,285 (2.7%)
architectural structure: 3,159,472 (4.4%)
occurrence: 3,898,674 (5.4%)
chemical compound: 1,188,724 (1.7%)
film: 294,370 (0.4%)
thoroughfare: 630,794 (0.9%)
astronomical object: 4,601,733 (6.4%)
Wikimedia list article: 404,454 (0.6%)
Wikimedia disambiguation page: 1,358,230 (1.9%)
Wikinews article: 195,900 (0.3%)
scholarly article: 22,574,314 (31.5%)
other P31/P279: 18,284,676 (25.5%)
no P31/P279: 3,973,469 (5.5%)

Google Knowledge Graph

"By March 2023, it had grown to 800 billion facts on 8 billion entities" (Wikipedia).



Wikidata and Google Knowledge Graph

WIKIDATA

Item | Discussion

Zsófia Báthory (Q250942)

Princess Consort of Transylvania (1629–1680)
Zsofia Bathory

edit

In more languages
Configure

Language	Label	Description	Also known as
English	Zsófia Báthory	Princess Consort of Transylvania (1629–1680)	Zsofia Bathory
German	Zsófia Báthory	Ehefrau von Georg II. Rákóczi, des Fürsten von Siebenbürgen (1629–1680)	
Alemannic	No label defined	No description defined	
French	Zsófia Báthory	(1629–1680)	

All entered languages

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page
Get shortened URL

Statements

instance of — human — 1 reference

image

Báthory Zsófia 1629.jpg
585 × 779; 155 KB

Google Knowledge Graph ID	/g/121258kx	edit
	0 references	+ add reference
		+ add value
Hungarian National Namespace person ID (new)	662639	edit
	0 references	+ add reference
		+ add value

WIKIDATA

Property | Discussion

Google Knowledge Graph ID (P2671)

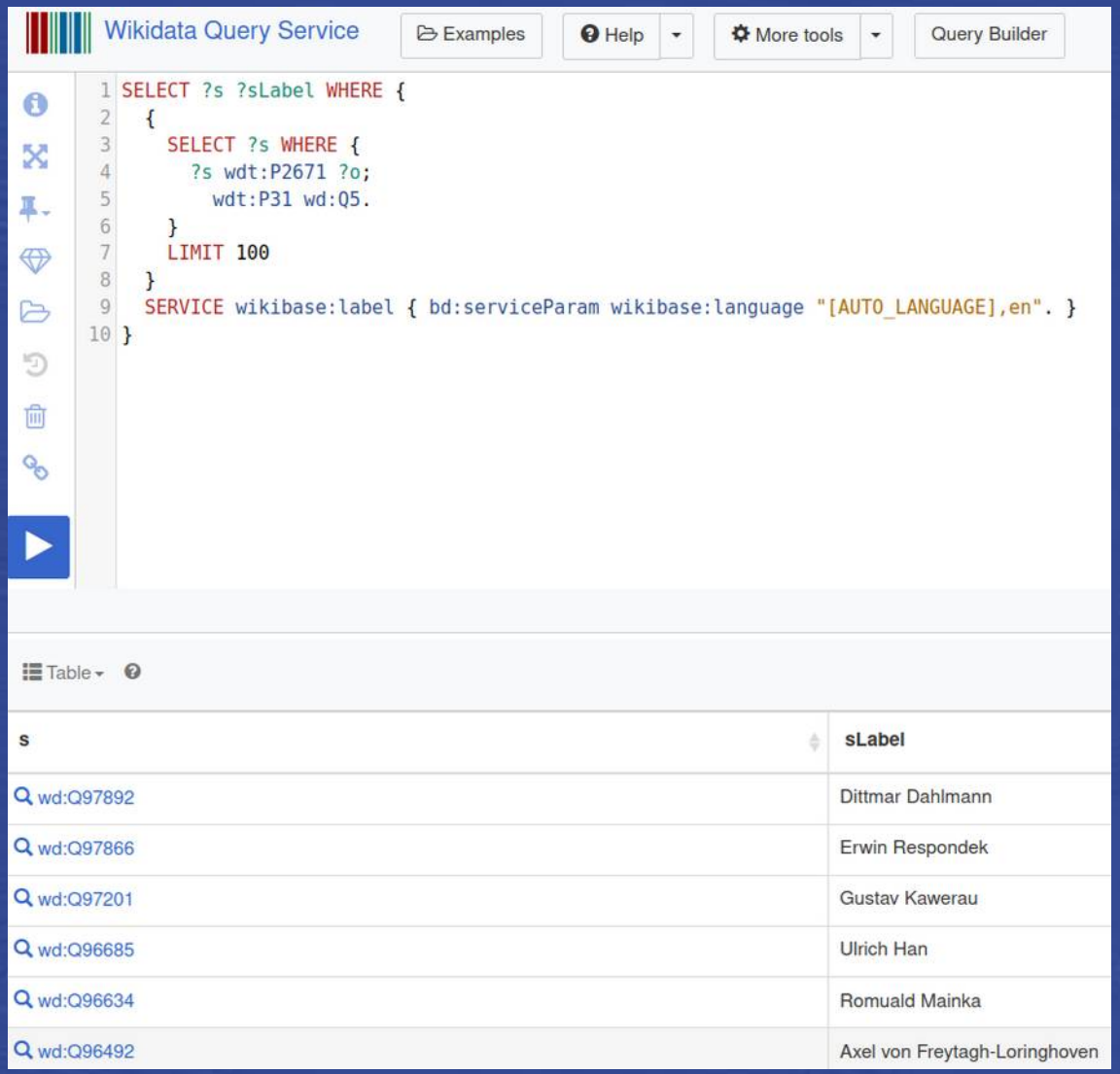
identifier for Google Knowledge Graph API, starting with "/g/". For IDs starting with "/m/", use Freebase ID (P646)

In more languages
Configure

Language	Label	Description	Also known as
English	Google Knowledge Graph ID	identifier for Google Knowledge Graph API, starting with "/g/". For IDs starting with "/m/", use Freebase ID (P646)	

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby

SPARQL : a query language for exploring LOD and the Semantic Web

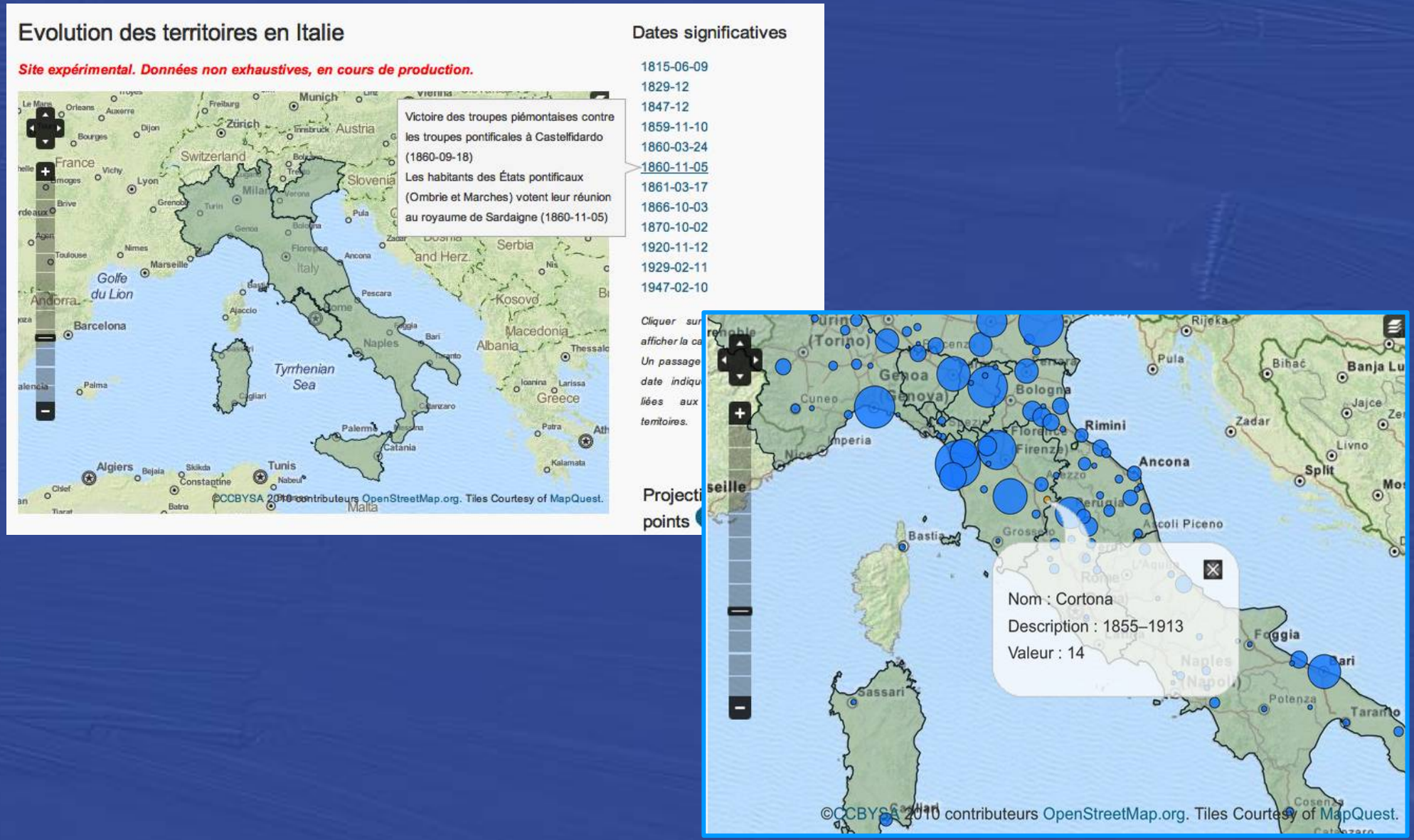


SPARQL – endpoint B3Kat

Bayerische Staatsbibliothek, Bibliotheksverbund Bayern, Kooperative Bibliotheksverbund Berlin-Brandenburg

https://lod.b3kat.de/doc/sparql-endpoint

Using library metadata published on the Semantic Web for historical research



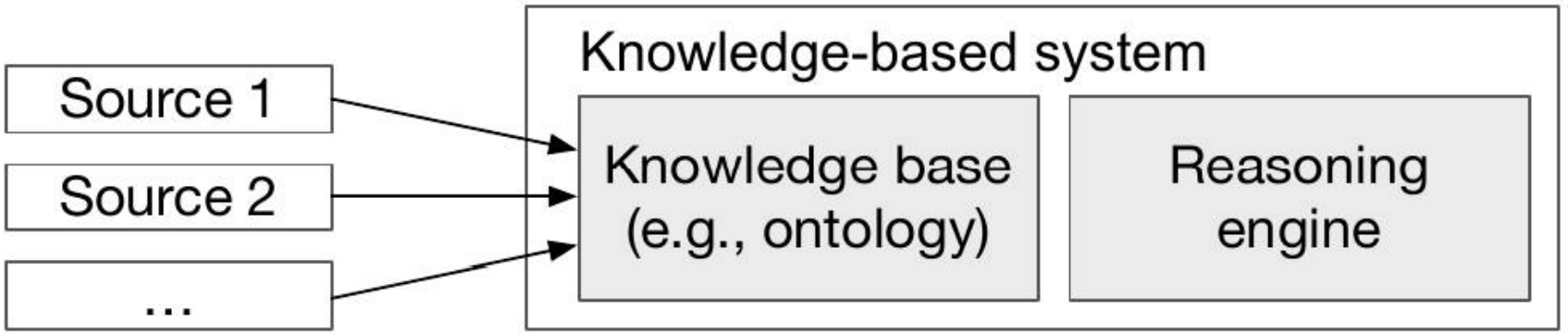


Figure 1: Architecture of a knowledge graph

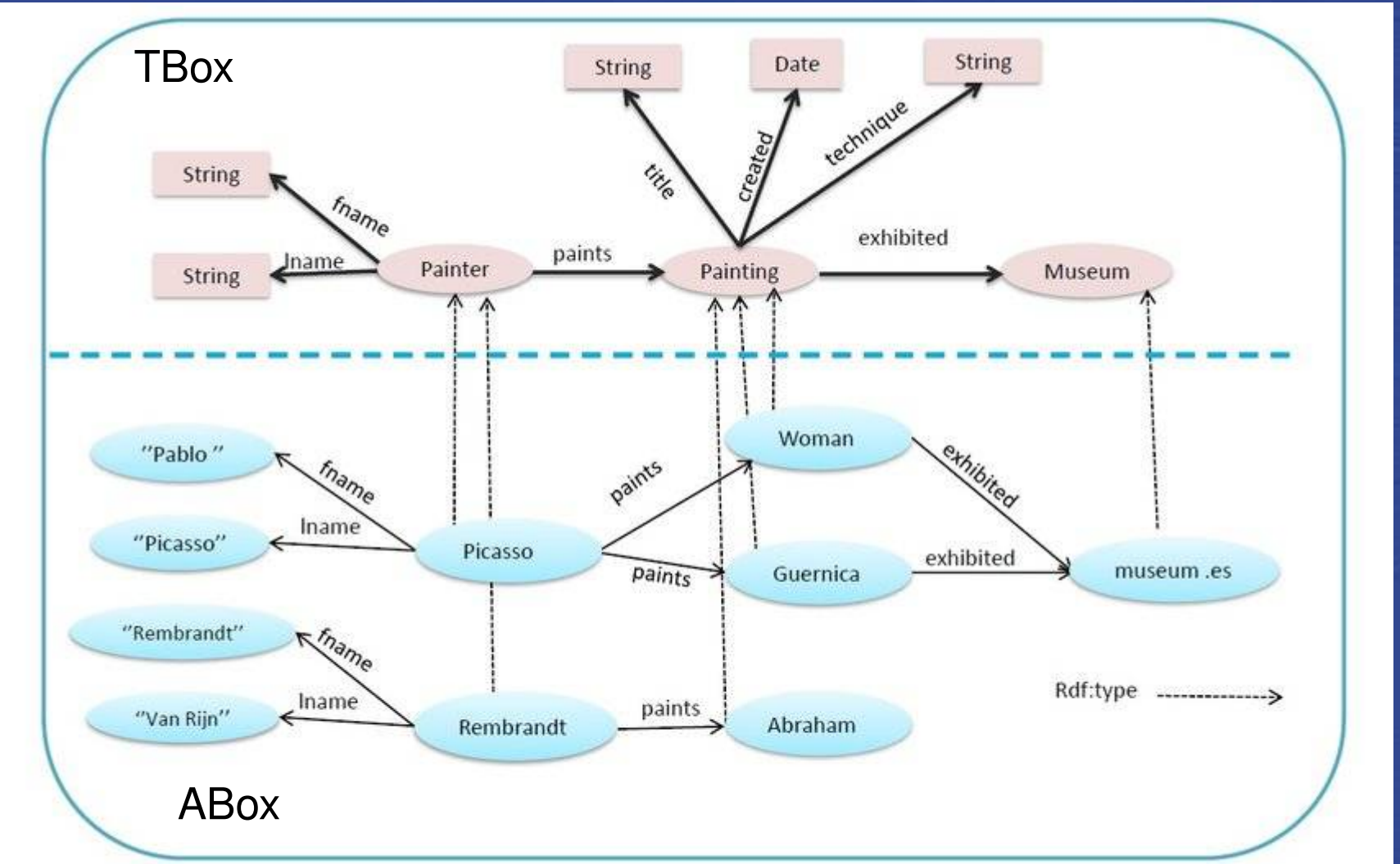
Ehrlinger, Lisa and Wolfram Wöß, 'Towards a Definition of Knowledge Graphs',

Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems ..., Leipzig, Germany, September 12-15, 2016, 2016 (https://ceur-ws.org/Vol-1695/paper4.pdf)

“An ontology is
a formal explicit specification
of a shared conceptualization
of a domain of interest”

- « Formality – ... a knowledge representation language that is based on the grounds of **formal semantics**. »
- « Consensus – … an agreement on a domain conceptualization among people in a community. »
- « Conceptuality – … in terms of conceptual symbols that can be intuitively grasped by humans, as they correspond to the elements in their **mental models**. »
- « Domain Specificity – ... limited to knowledge about a particular **domain of interest**. »

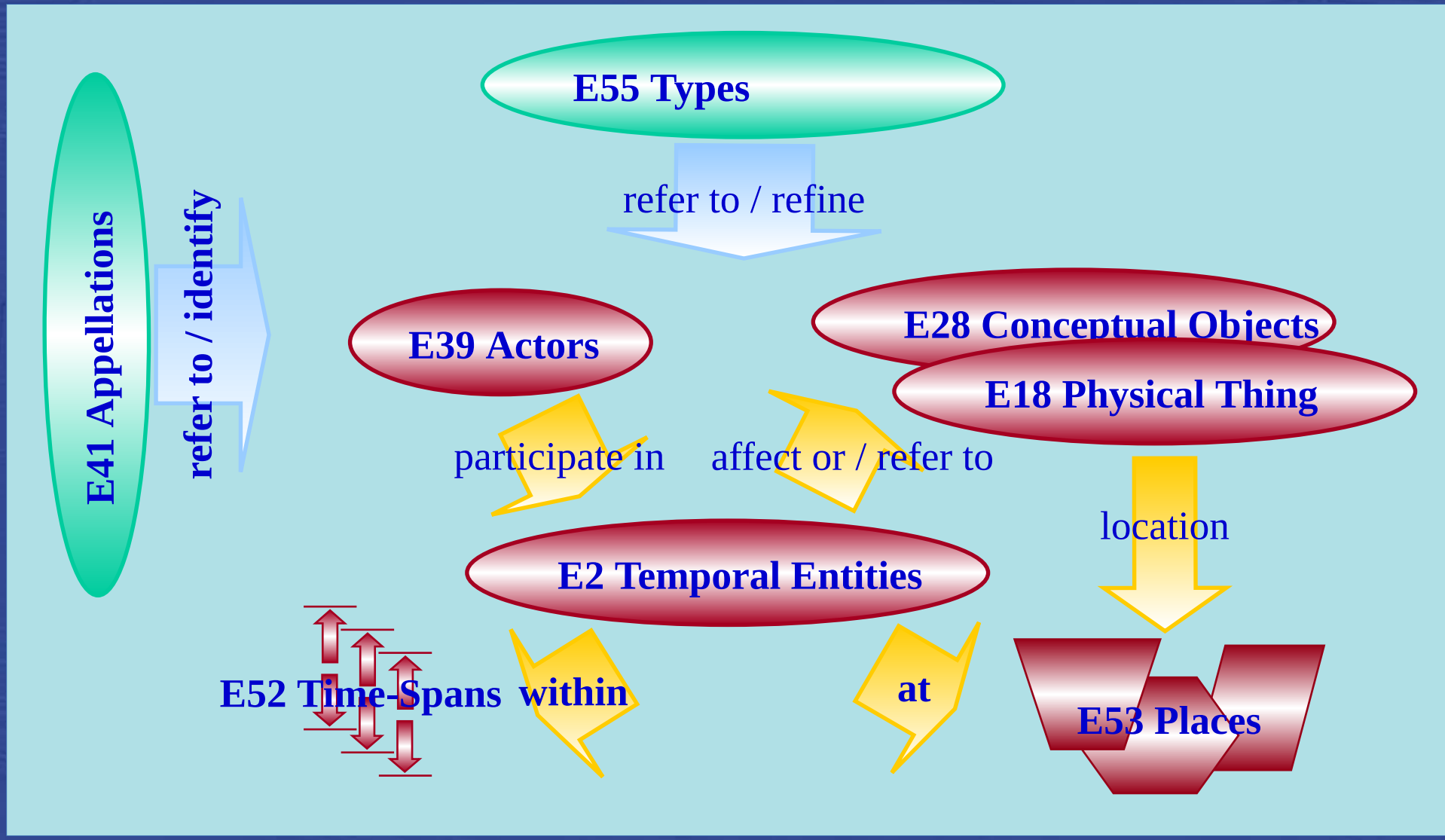
[Domingue et al. 2011, p. 510-511]



Dimitris Kotzinos

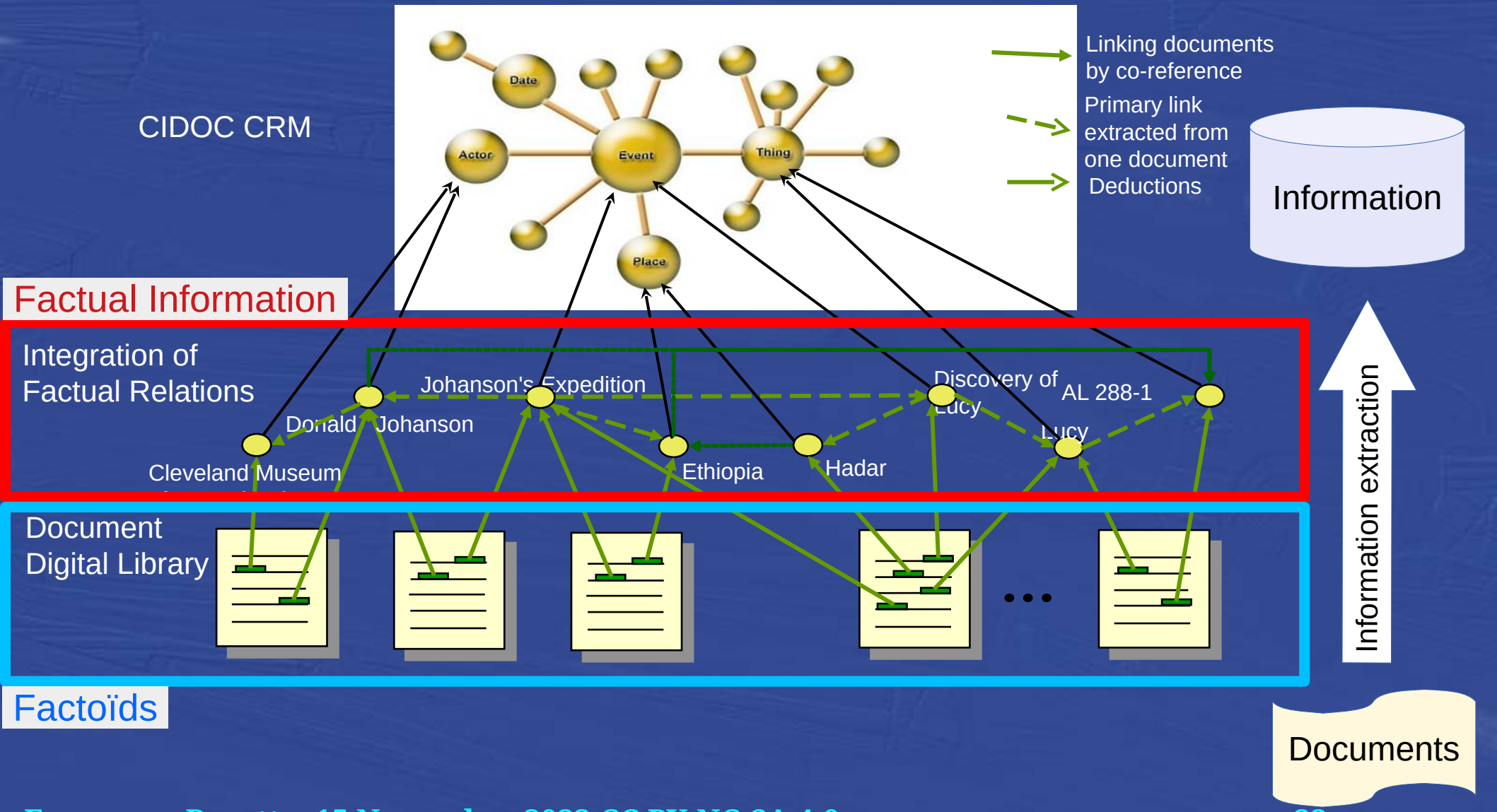
RDFS – OWL-DL

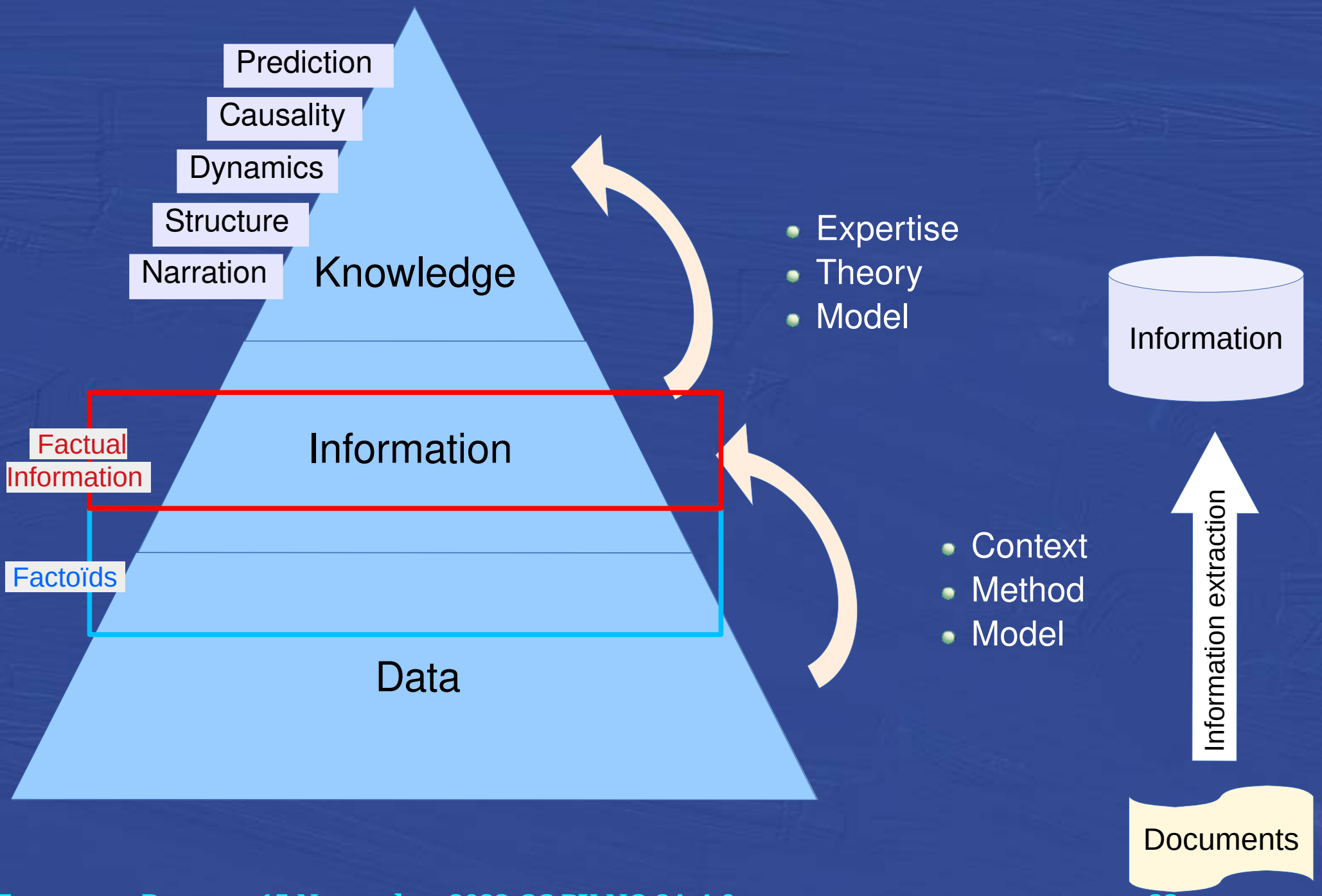
The CIDOC CRM (ISO21127:2006)
A semantic framework that provides *interoperability* between different sources of **cultural heritage information**

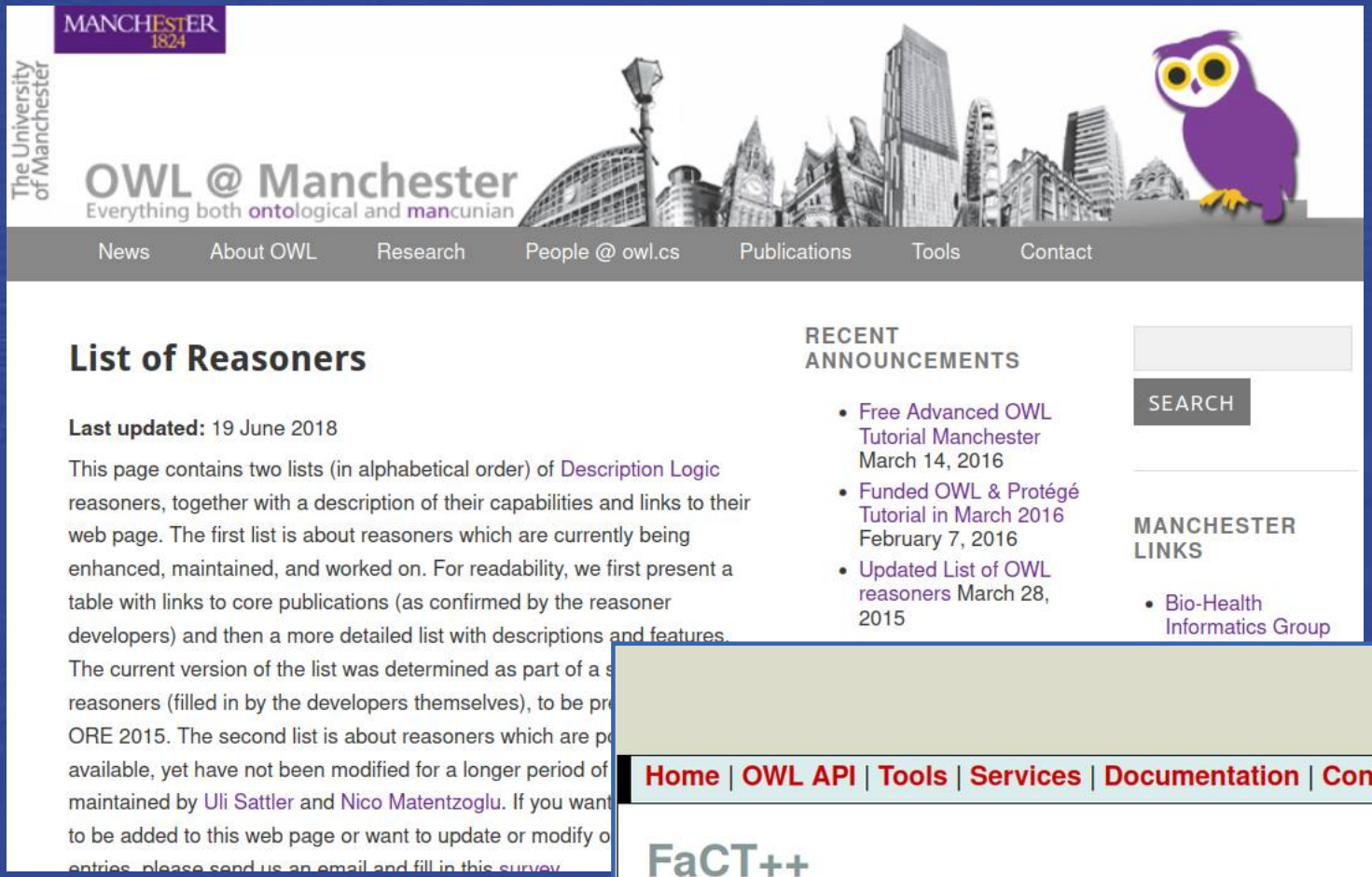


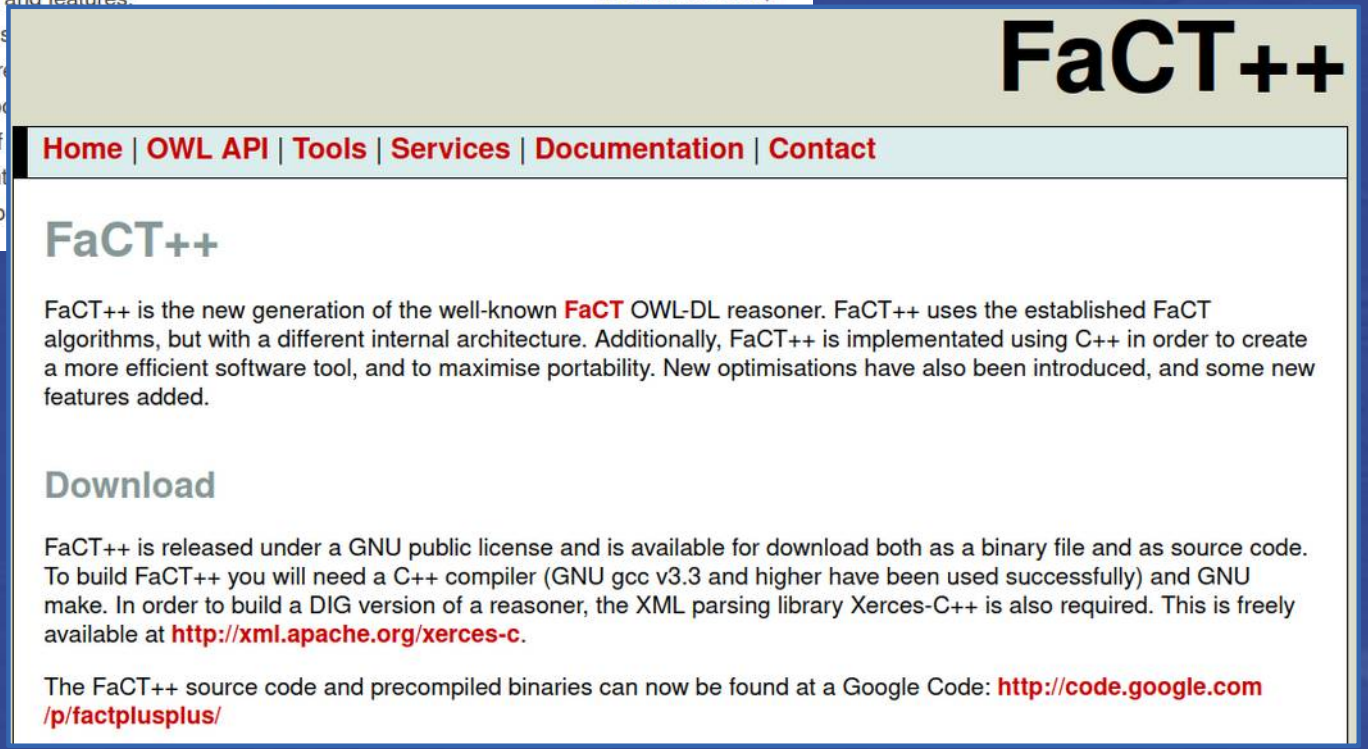
Stephen Stead (2008)

Integration of information extracted from documents using the CIDOC CRM




Prediction
Causality
Dynamics
Structure
Narration
Knowledge
Information
Factual Information
Factoïds
Data
Expertise
Theory
Model
Context
Method
Model
Information
Information extraction
Documents


MANCHESTER 1824
The University of Manchester
OWL @ Manchester
Everything both ontological and mancunian
News
About OWL
Research
People @ owl.cs
Publications
Tools
Contact
List of Reasoners
Last updated: 19 June 2018
This page contains two lists (in alphabetical order) of Description Logic reasoners, together with a description of their capabilities and links to their web page. The first list is about reasoners which are currently being enhanced, maintained, and worked on. For readability, we first present a table with links to core publications (as confirmed by the reasoner developers) and then a more detailed list with descriptions and features.
The current version of the list was determined as part of a s
reasoners (filled in by the developers themselves), to be pr
ORE 2015. The second list is about reasoners which are p
available, yet have not been modified for a longer period of
maintained by Uli Sattler and Nico Matentzoglu. If you want
to be added to this web page or want to update or modify o
entries, please send us an email and fill in this survey
RECENT ANNOUNCEMENTS
Free Advanced OWL Tutorial Manchester March 14, 2016
Funded OWL & Protégé Tutorial in March 2016 February 7, 2016
Updated List of OWL reasoners March 28, 2015
SEARCH
MANCHESTER LINKS
Bio-Health Informatics Group


FaCT++
Home | OWL API | Tools | Services | Documentation | Contact
FaCT++
FaCT++ is the new generation of the well-known FaCT OWL-DL reasoner. FaCT++ uses the established FaCT algorithms, but with a different internal architecture. Additionally, FaCT++ is implementated using C++ in order to create a more efficient software tool, and to maximise portability. New optimisations have also been introduced, and some new features added.
Download
FaCT++ is released under a GNU public license and is available for download both as a binary file and as source code. To build FaCT++ you will need a C++ compiler (GNU gcc v3.3 and higher have been used successfully) and GNU make. In order to build a DIG version of a reasoner, the XML parsing library Xerces-C++ is also required. This is freely available at http://xml.apache.org/xerces-c.
The FaCT++ source code and precompiled binaries can now be found at a Google Code: http://code.google.com /p/factplusplus/

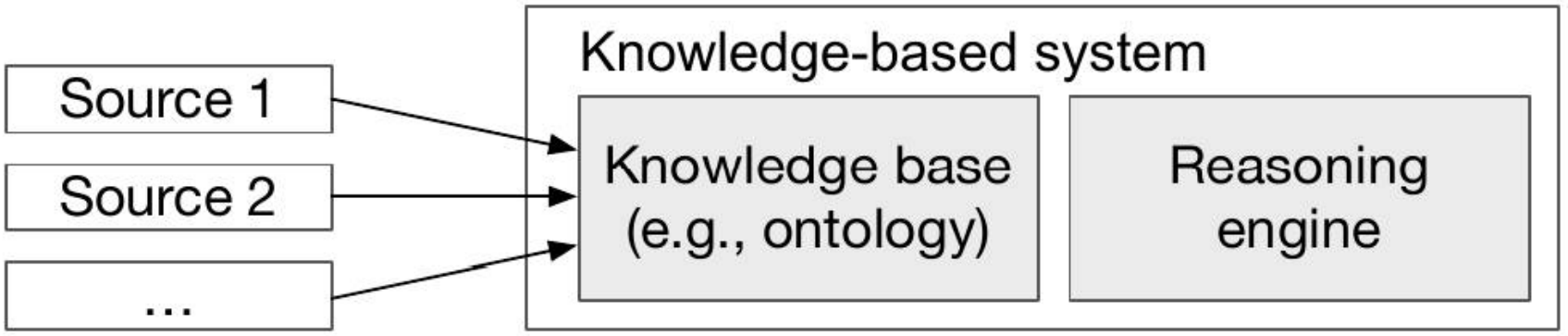
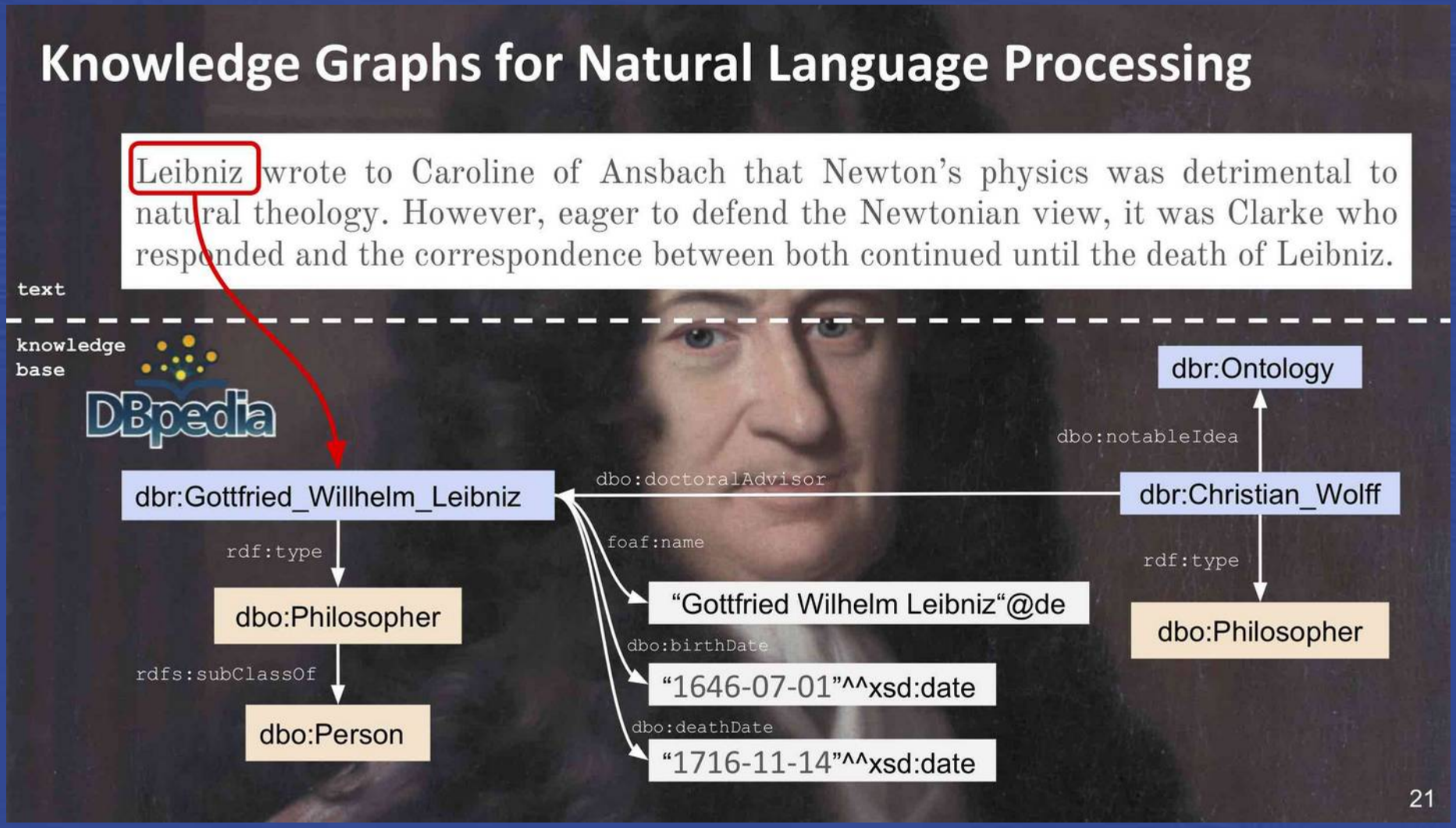


Figure 1: Architecture of a knowledge graph

Ehrlinger, Lisa and Wolfram Wöß, 'Towards a Definition of Knowledge Graphs',

Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems ..., Leipzig, Germany, September 12-15, 2016, 2016 (https://ceur-ws.org/Vol-1695/paper4.pdf)



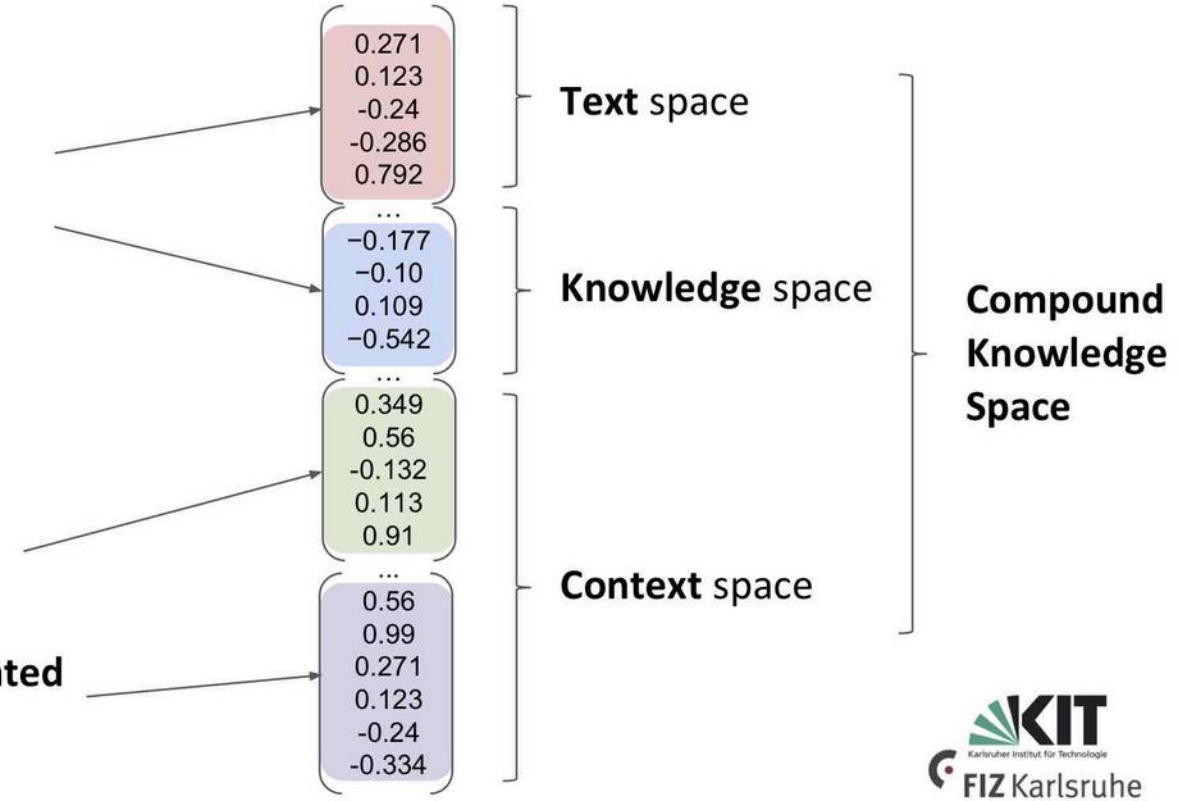
Karlsruher Institut für Technologie (29. Novembre 2017) – Antrittsvorlesung von

Prof. Dr. Harald Sack

Combining Semantics and Deep Learning for Intelligent Information Services

Combined Feature Embeddings for a Compound Knowledge Space

- Various feature vectors
  - **Word** embeddings
  - **Knowledge Graph** embeddings
    - Instances
    - Ontologies
  - Embeddings for **semantically enriched texts**
  - **Metadata** and **aggregated features**

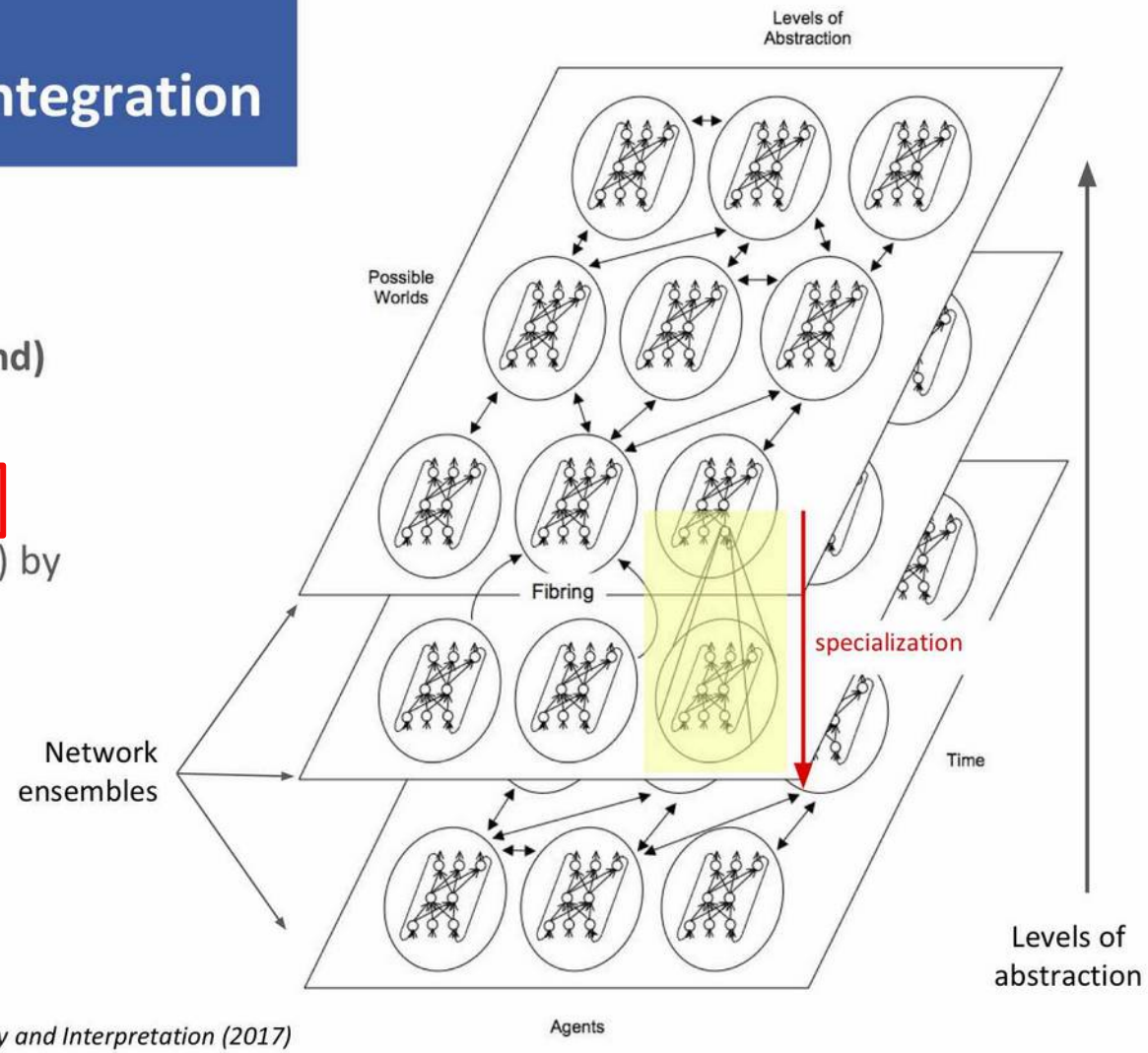


Antrittsvorlesung von Prof. Dr. Harald Sack (29. Novembre 2017)

Towards Neuro-Symbolic Integration

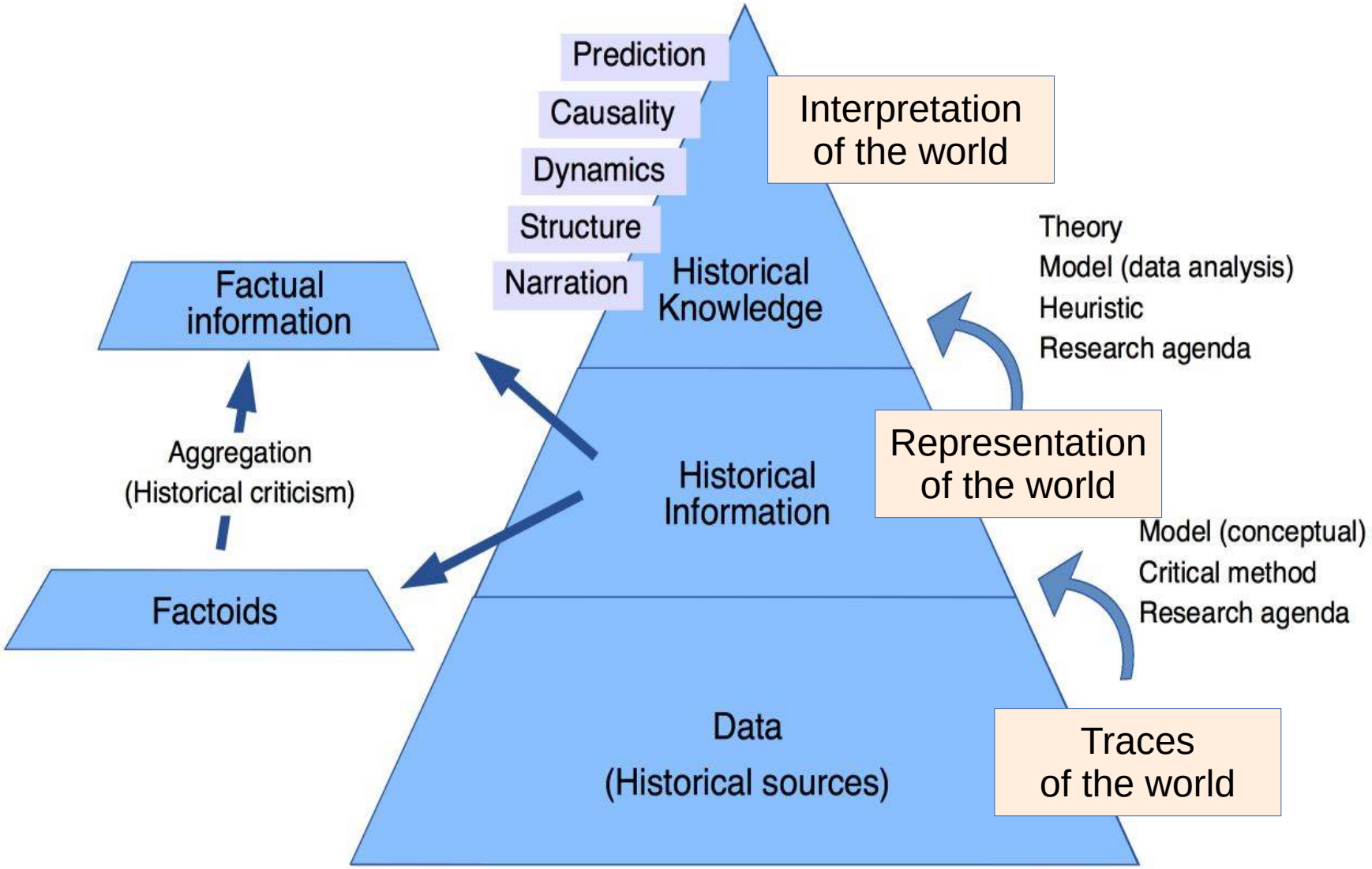
Neuro-Symbolic Systems

1. Translation of **symbolic (background) knowledge into the network**
2. **Learning of additional knowledge** from examples (and generalisation) by the network
3. **Executing the network** (i.e. reasoning), and
4. **Symbolic knowledge extraction** from the network.



33 *Besold et al.: Neural-Symbolic Learning and Reasoning: A Survey and Interpretation (2017)*

Antrittsvorlesung von Prof. Dr. Harald Sack (29. Novembre 2017)



Collaboration Université
de Berne et LARHRA
CNRS/Université de Lyon