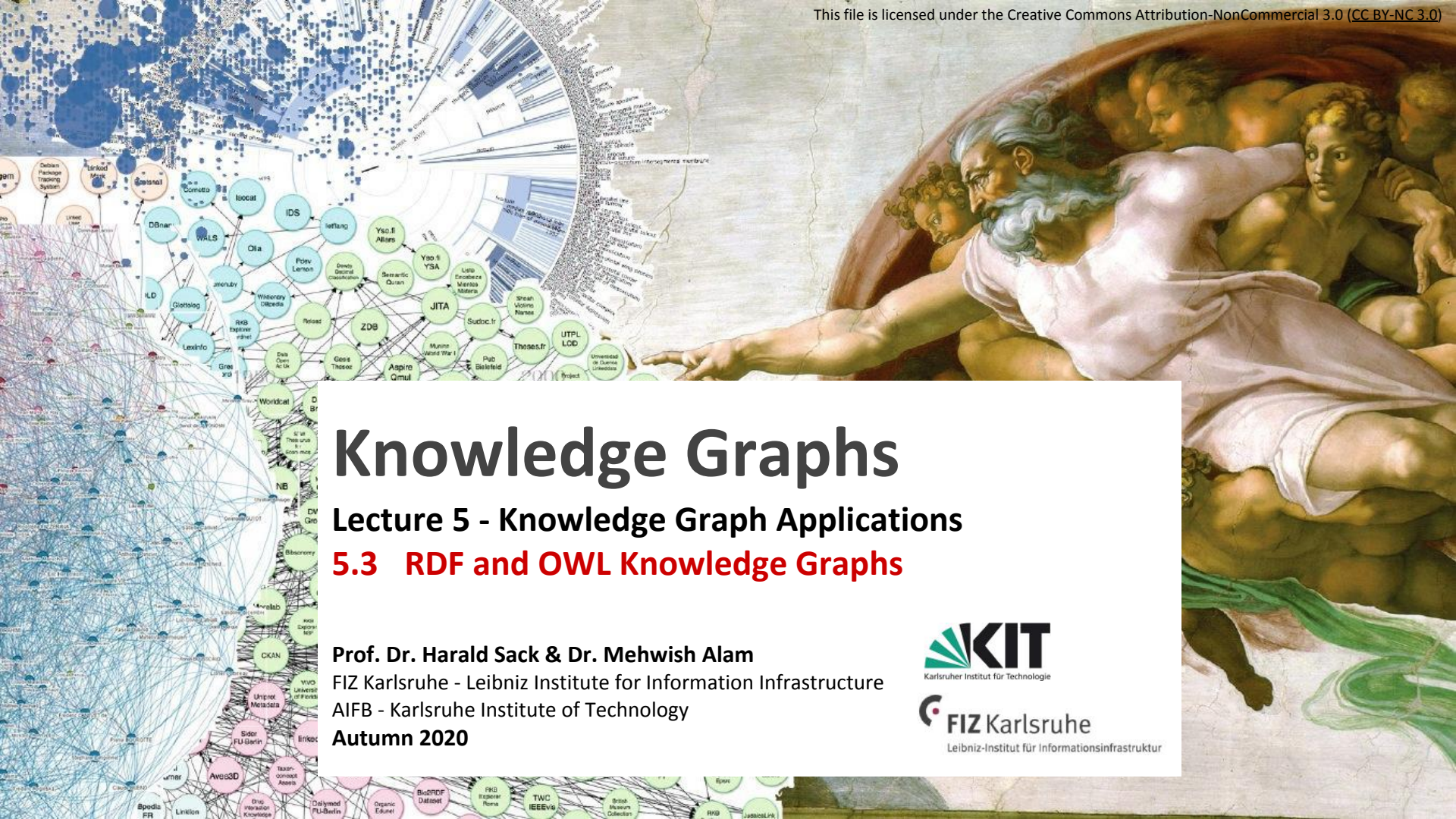


Knowledge Graphs

Lecture 5 - Knowledge Graph Applications

5.3 RDF and OWL Knowledge Graphs

Prof. Dr. Harald Sack & Dr. Mehwish Alam
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology
Autumn 2020

Karlsruher Institut für Technologie

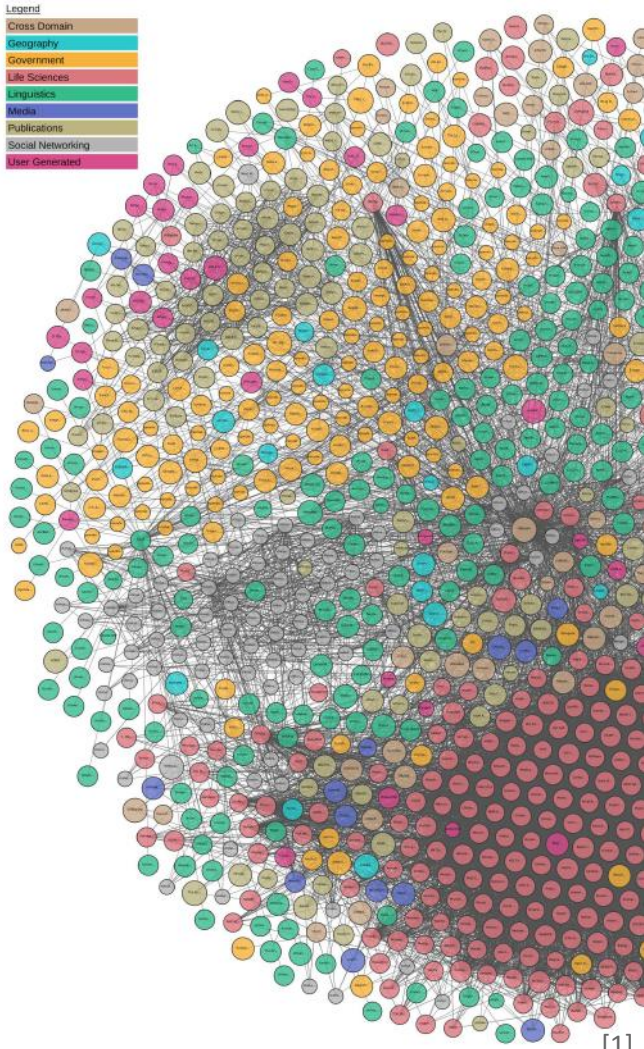
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Knowledge Graphs

Lecture 5: Knowledge Graph Applications

Creating Knowledge Graphs

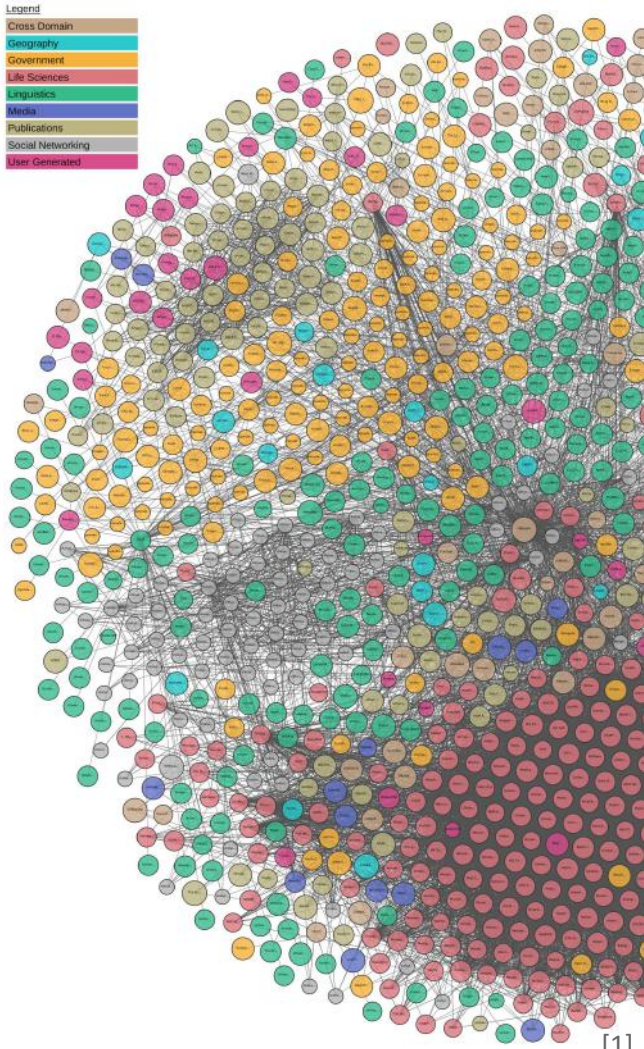
- **Curated approaches:**
  - Triples are created manually by a closed group of experts.
- **Collaborative approaches:**
  - Triples are created manually by an open group of volunteers.
- **Automated semi-structured approaches:**
  - Triples are extracted automatically from semi-structured text via hand-crafted rules, learned rules, or regular expressions.
- **Automated unstructured approaches:**
  - Triples are extracted automatically from unstructured text via machine learning and natural language processing techniques.
- **Linking existing datasets:**
  - Different dataset are connected using linked data.



[1]

Creating Knowledge Graphs

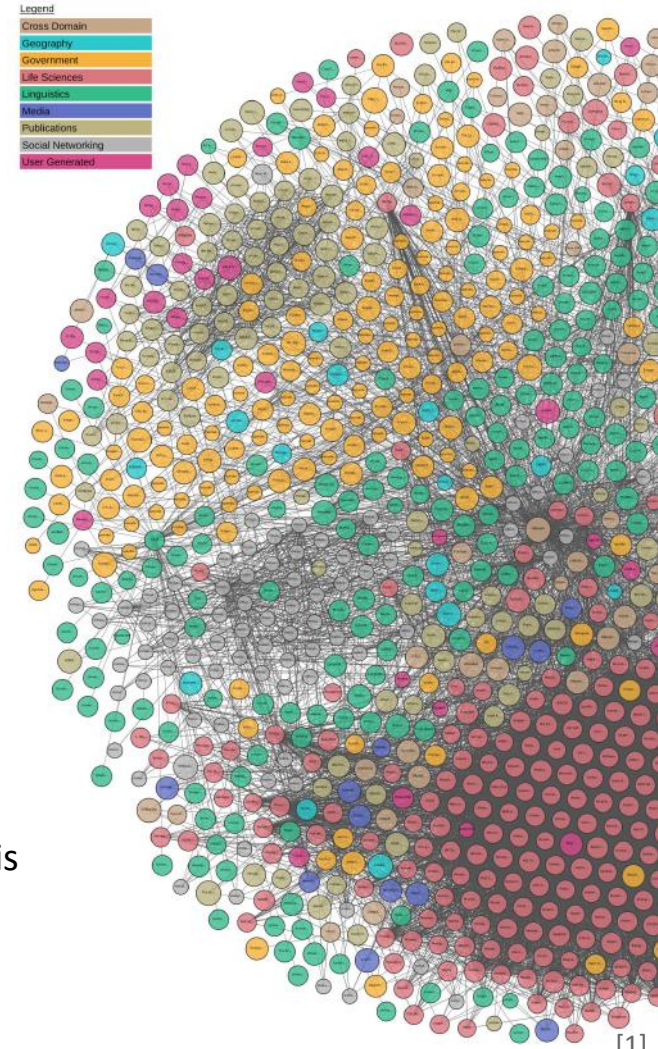
- **Curated approaches:**
  - Triples are created manually by a closed group of experts.
  - Highly accurate
  - Don’t scale
  - Examples:
    - Cyc/OpenCyc
    - WordNet
    - UMLS (United Medical Language System)
    - SNOMED CT



[1]

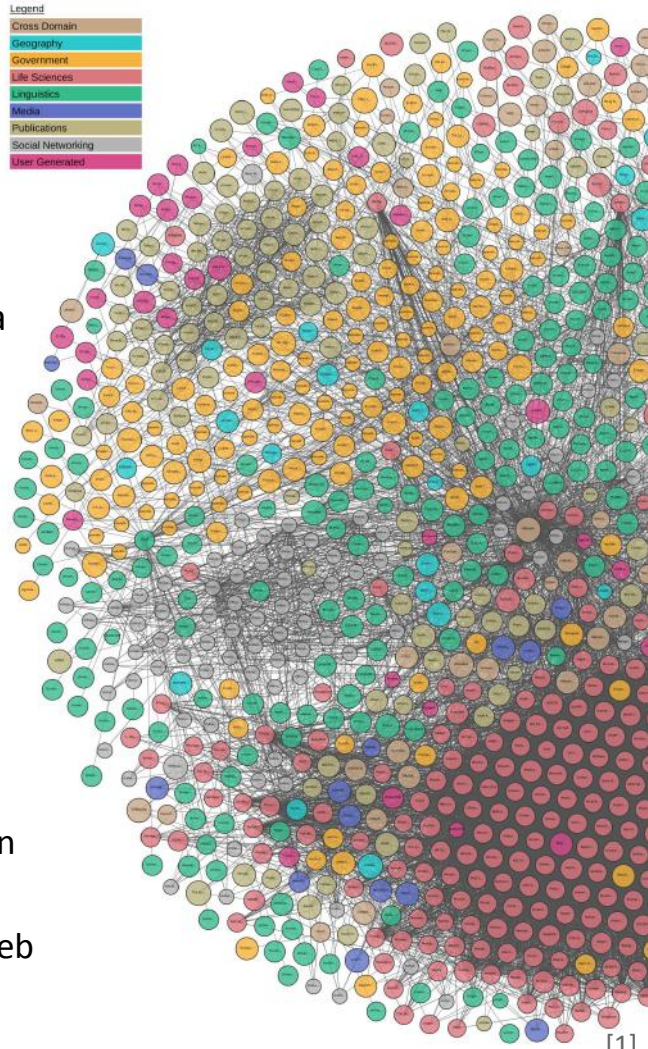
Creating Knowledge Graphs

- **Collaborative approaches:**
  - Triples are created manually by an open group of volunteers.
  - Better scaling
  - Examples:
    - Wikidata
    - Freebase
  - Issues:
    - Incompleteness: the (mandatory) place of birth attribute is missing for 71% of all people included in Freebase
    - The growth of Wikipedia/Wikidata is slowing down



[1]

Creating Knowledge Graphs

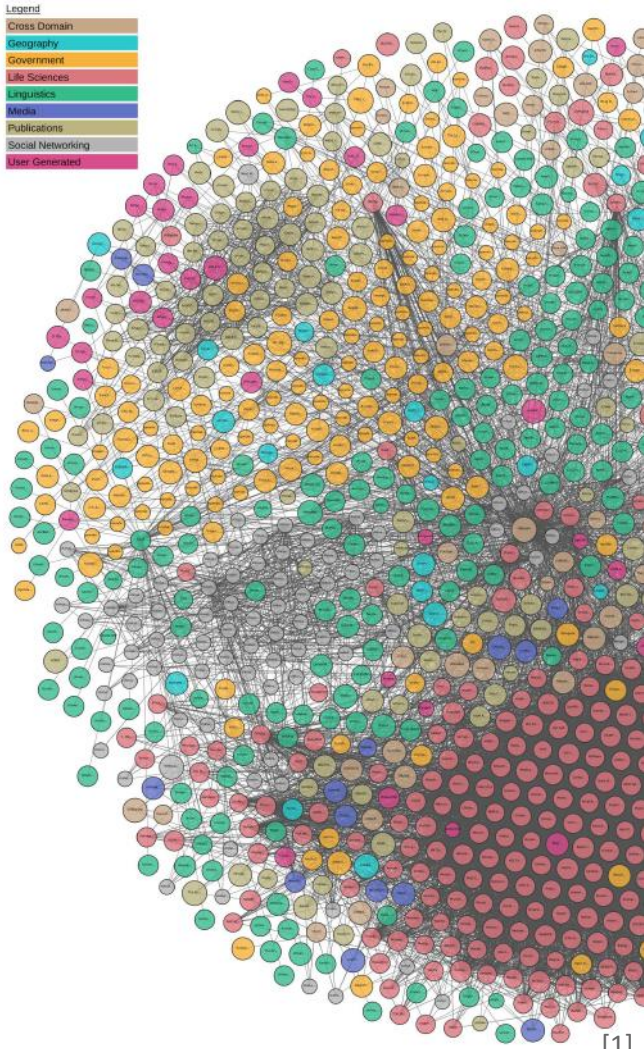


- **Automated semi-structured approaches:**
  - Triples are extracted automatically from semi-structured text via hand-crafted rules, learned rules, or regular expressions.
  - Exploits semi-structured data such as Wikipedia infoboxes
  - Scale and accuracy
  - Examples:
    - YAGO
    - DBpedia
    - Freebase
  - Accuracy of YAGO2 estimated at over 95% through manual inspection
  - Accuracy of Freebase estimated to be 99%
  - Coverage of only a small fraction of the information stored on the Web

[1]

Creating Knowledge Graphs

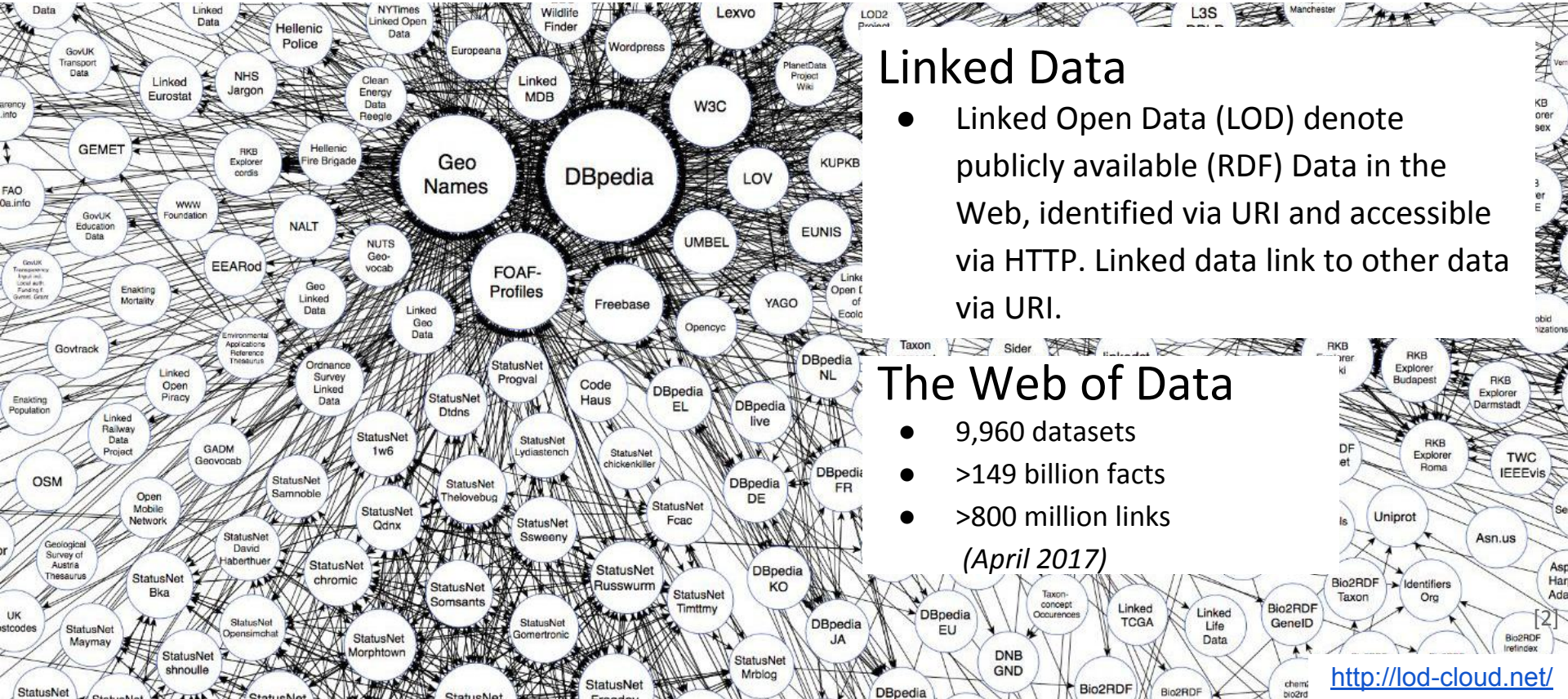
- **Automated unstructured approaches:**
  - Triples are extracted automatically from unstructured text via machine learning and natural language processing techniques.
  - Facts extracted from the natural language text of Web pages.
  - Examples:
    - NELL, Knowledge Vault, PATTY, PROSPERA,
    - DeepDive/Elementary,
    - ReVerb, OLLIE, PRISMATIC
  - **Noise**: can be reduced by using the knowledge from existing, high-quality repositories.



[1]

The Web of Linked Data - A Web of Knowledge Graphs

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure



Linked Data

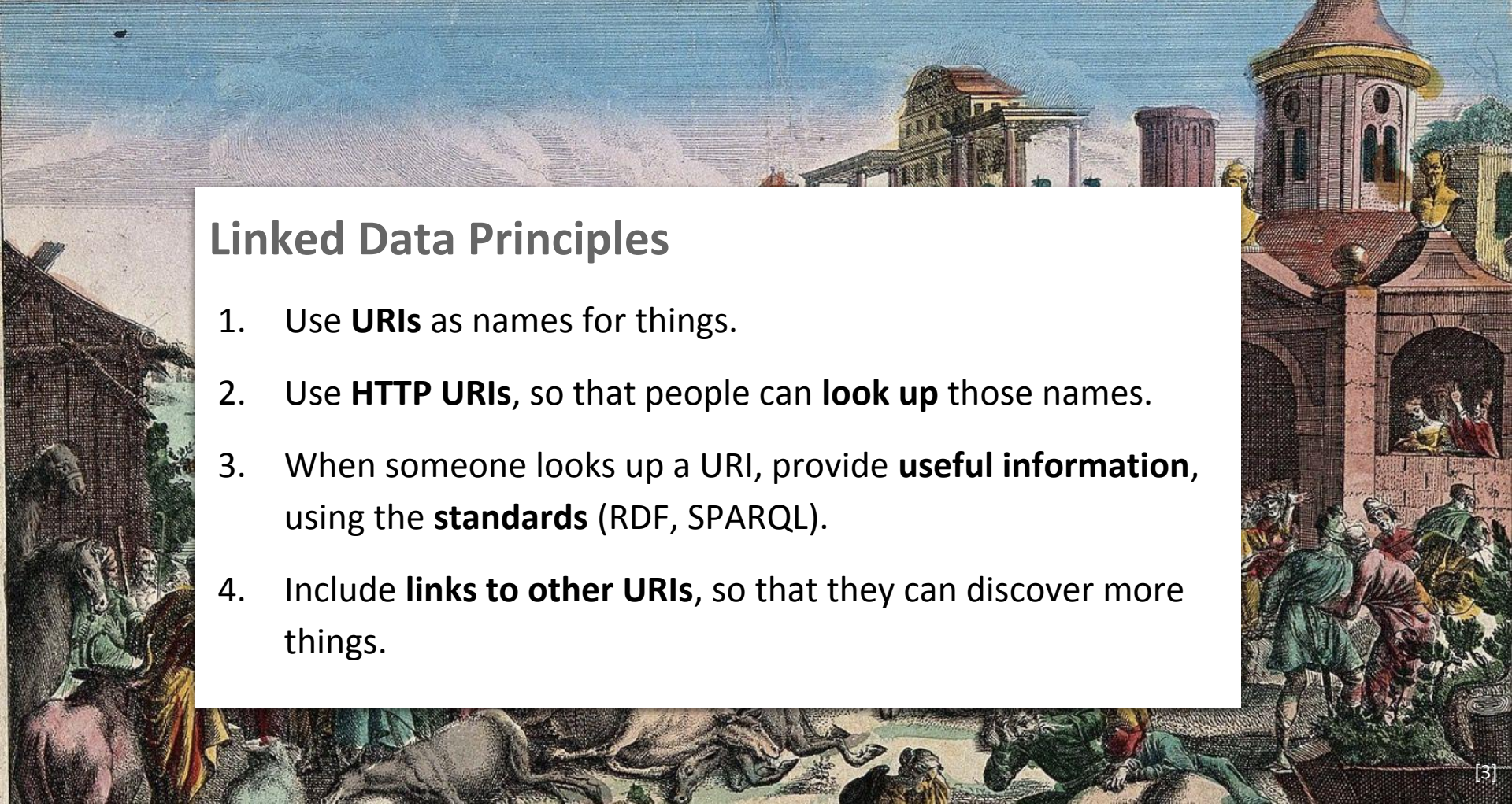
- Linked Open Data (LOD) denote publicly available (RDF) Data in the Web, identified via URI and accessible via HTTP. Linked data link to other data via URI.

The Web of Data

- 9,960 datasets
- >149 billion facts
- >800 million links

(April 2017)

http://lod-cloud.net/



Linked Data Principles

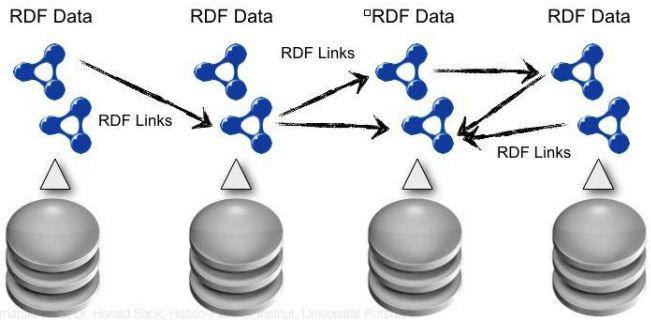
1. Use **URIs** as names for things.
2. Use **HTTP URIs**, so that people can **look up** those names.
3. When someone looks up a URI, provide **useful information**, using the **standards** (RDF, SPARQL).
4. Include **links to other URIs**, so that they can discover more things.

[3]

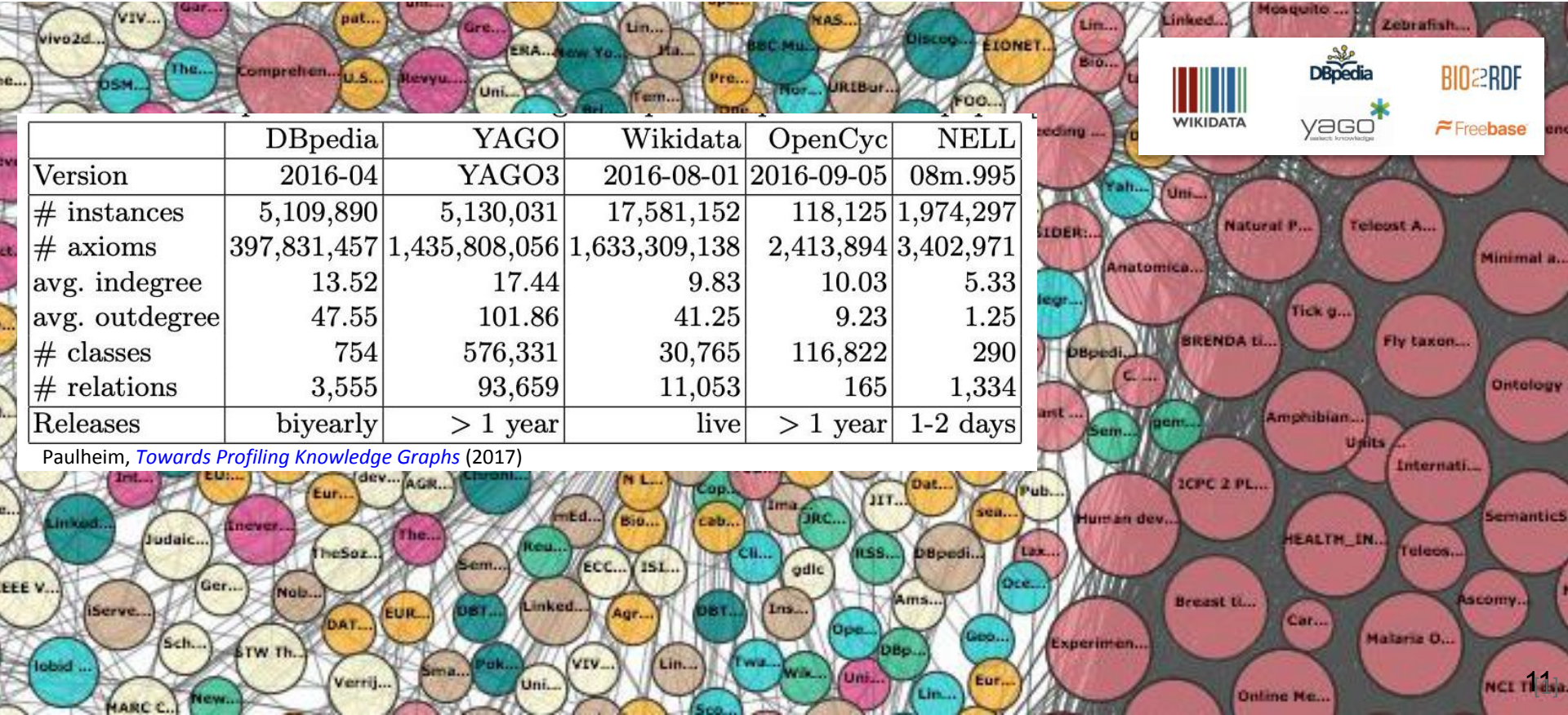
(Tim Berners-Lee, Linked Data, 2006, http://www.w3.org/DesignIssues/LinkedData.html)

Advantages of Linked Open Data vs. APIs

- **Simple** and **generic API** for various heterogeneous data sources enables **simple reuse** and **data sharing** among applications
- **RDF Data model** guarantees (simple) **extensibility**
- **Transport via http**, standard Port 80, prevents firewall adaption
- **Ontologies** enable **meaningful connections** between data sources
- **Reasoning** over Linked Data enables to **generate new knowledge**, i.e. inference from implicit to explicit knowledge



Popular (Open) Knowledge Graphs



	DBpedia	YAGO	Wikidata	OpenCyc	NELL
Version	2016-04	YAGO3	2016-08-01	2016-09-05	08m.995
# instances	5,109,890	5,130,031	17,581,152	118,125	1,974,297
# axioms	397,831,457	1,435,808,056	1,633,309,138	2,413,894	3,402,971
avg. indegree	13.52	17.44	9.83	10.03	5.33
avg. outdegree	47.55	101.86	41.25	9.23	1.25
# classes	754	576,331	30,765	116,822	290
# relations	3,555	93,659	11,053	165	1,334
Releases	biyearly	> 1 year	live	> 1 year	1-2 days

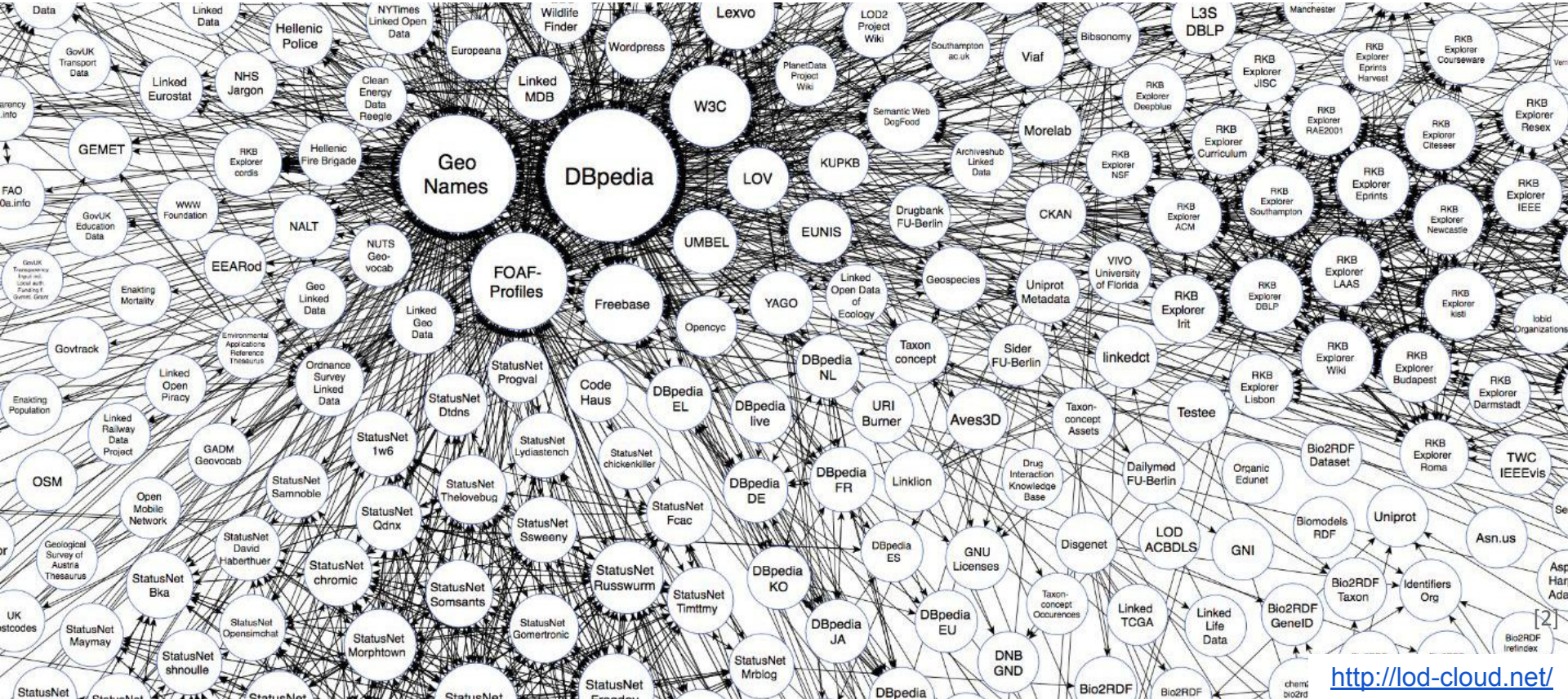
Paulheim, *Towards Profiling Knowledge Graphs* (2017)

Popular (Proprietary) Knowledge Graphs

	Data model	Size of the graph	Development stage
Microsoft	The types of entities, relations, and attributes in the graph are defined in an ontology.	~2 billion primary entities, ~55 billion facts	Actively used in products
Google	Strongly typed entities, relations with domain and range inference	1 billion entities, 70 billion assertions	Actively used in products
Facebook	All of the attributes and relations are structured and strongly typed, and optionally indexed to enable efficient retrieval, search, and traversal.	~50 million primary entities, ~500 million assertions	Actively used in products
eBay	Entities and relation, well-structured and strongly typed	Expect around 100 million products, >1 billion triples	Early stages of development and deployment
IBM	Entities and relations with evidence information associated with them.	Various sizes. Proven on scales documents >100 million, relationships >5 billion, entities >100 million	Actively used in products and by clients

Noy et al,
Industry-scale Knowledge Graphs: Lessons and Challenges (2019)

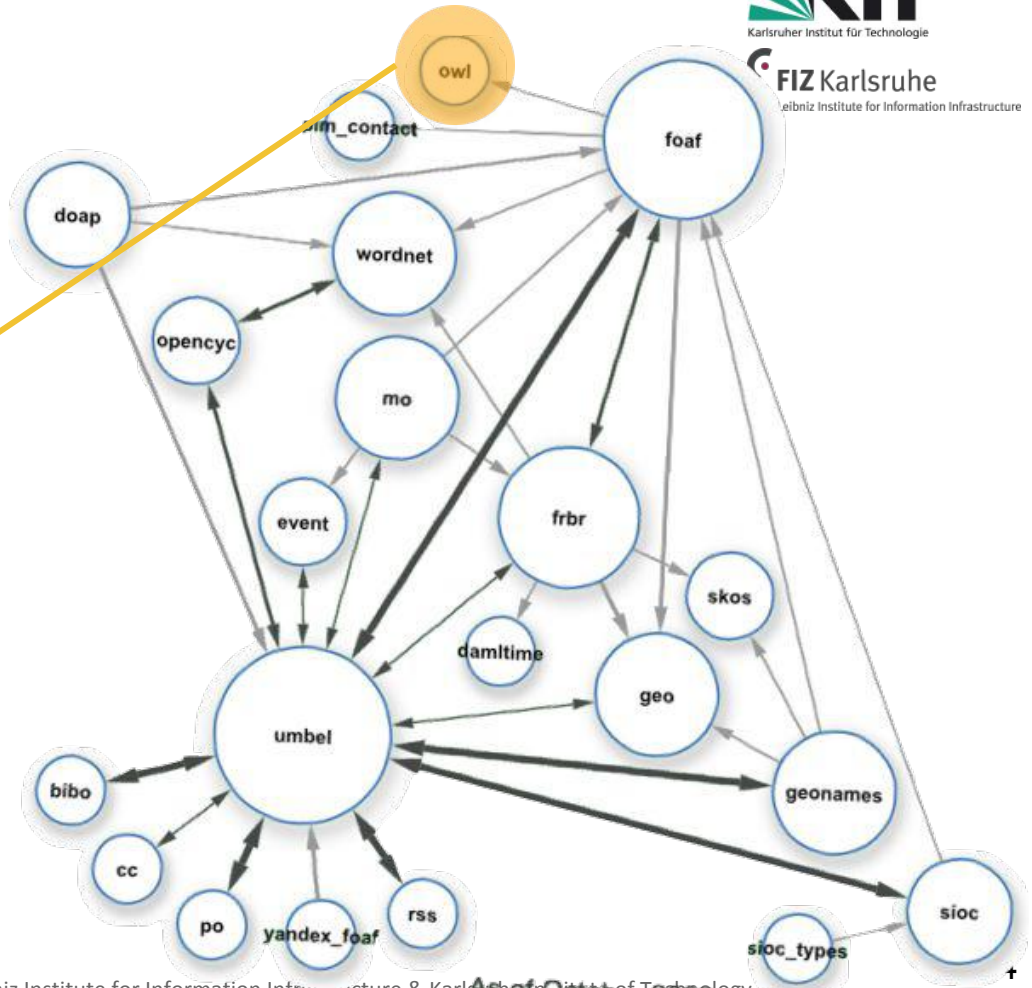
What keeps the Web of Linked Data together?



Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

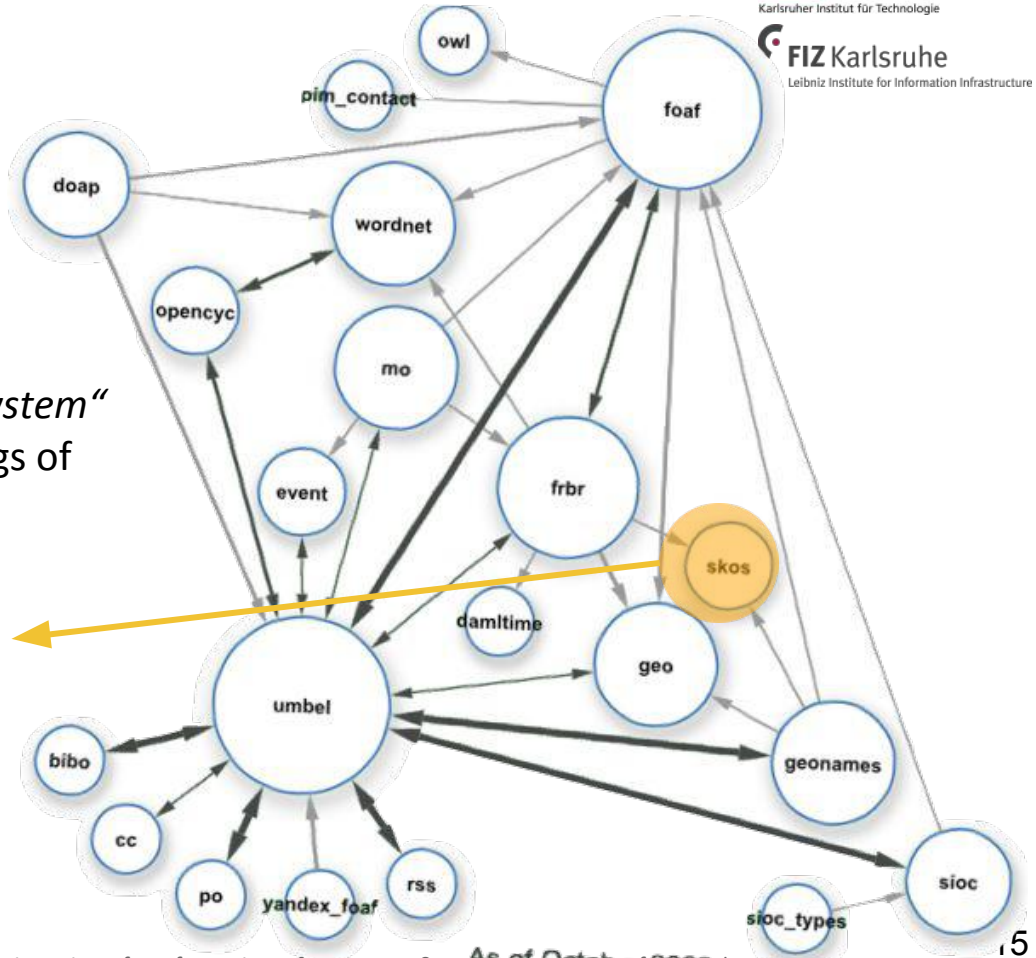
- **OWL**
  `owl:sameAs` connects identical individuals
  `owl:equivalentClass` connects equivalent classes



Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

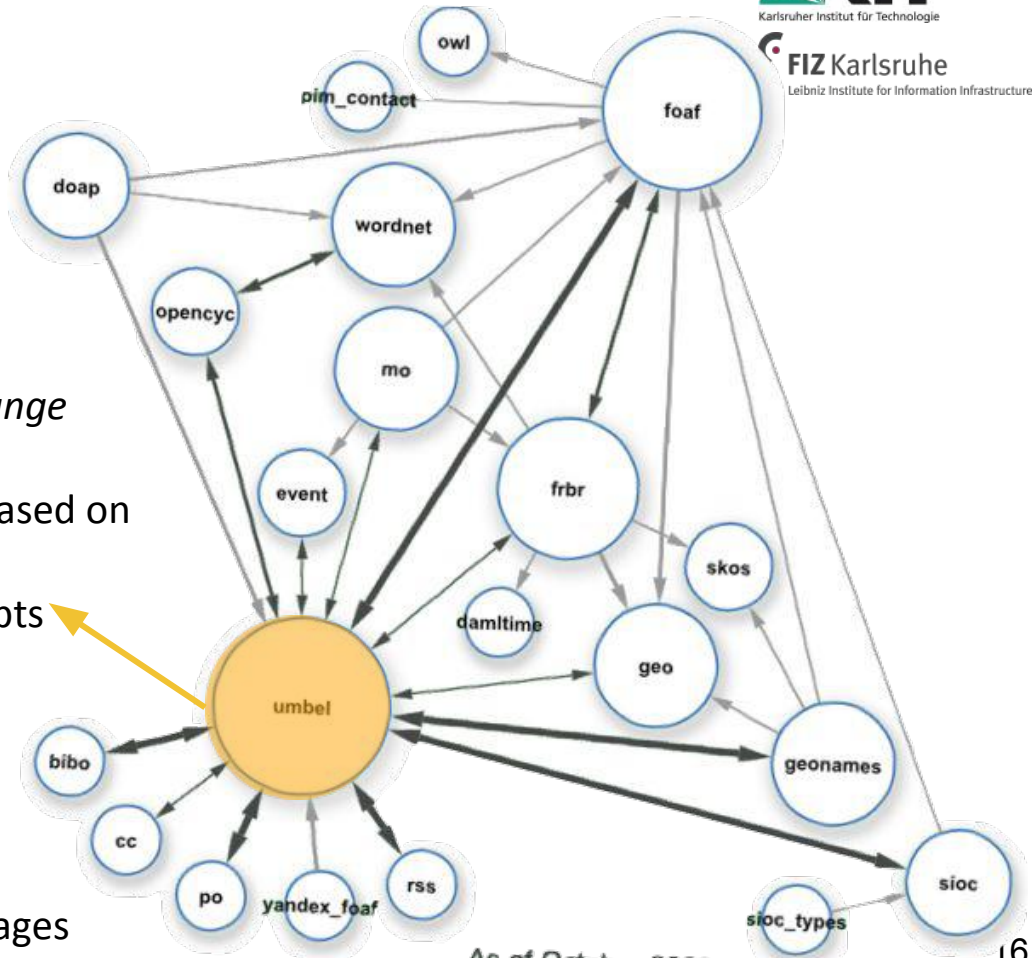
- **SKOS**
  - *„Simple Knowledge Organization System“*
  - Applied for definitions and mappings of vocabularies and ontologies
    - `skos:Concept`
    - `skos:narrower`
    - `skos:broader`
    - `skos:related`
    - `skos:exactMatch`, `skos:narrowMatch`, `skos:broadMatch`, `skos:relatedMatch`

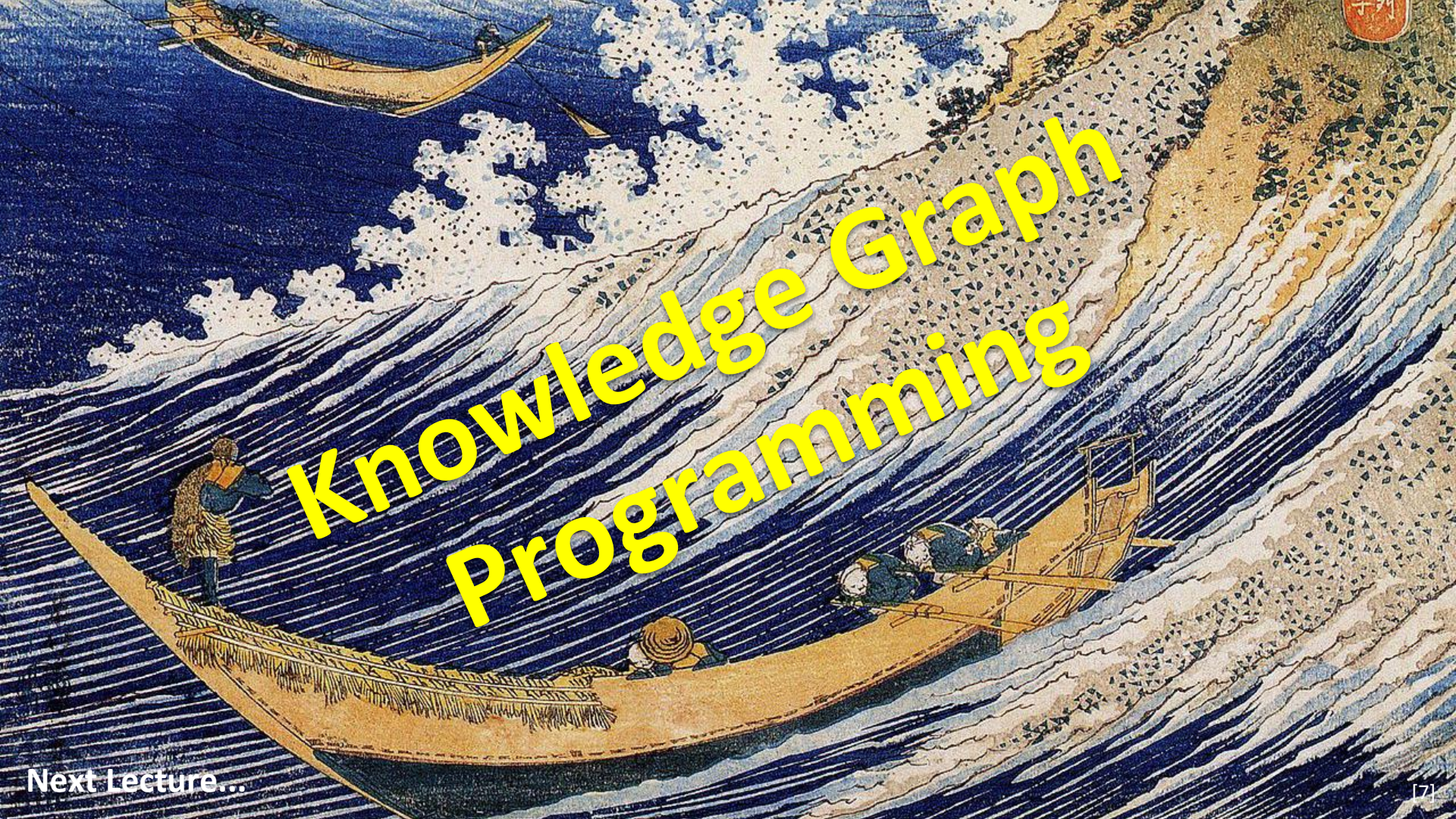


Linked Data Ontologies

- Ontologies hold the Linked Data Cloud together

- **UMBEL**
  - *„Upper Mapping and Binding Exchange Layer“*
  - Subset of OpenCyc as RDF Triples based on SKOS and OWL
  - Upper Ontology with 28.000 concepts (`skos:Concept`)
  - 46.000 Mappings into DBpedia, geonames, e.a. (`owl:equivalentClass`, `rdfs:subClassOf`)
  - Links to more than 2m Wikipedia pages





Next Lecture...

Picture References:

- [1] John P. McCrae, The Linked Open Data Cloud, [CC-BY-4.0] https://lod-cloud.net/
- [2] The Linked Open Data Cloud, 2014 version, [CC-BY-4.0], https://lod-cloud.net/versions/2014-08-30/lod-cloud.png
- [3] The fifth plague of Egypt, cattle dying. The Wellcome Collection, [CC-BY-4.0] https://wellcomecollection.org/works/m8g9bptf
- [4] Hokusai, « Choshi dans la province de Soshu » , série : Mille images de la mer. Vers 1832-34. Estampe nishiki-e, [Public Domain]], https://commons.wikimedia.org/wiki/File:Hokusai_1760-1849_Ocean_waves.jpg