JAGIELLONIAN UNIVERSITY
IN KRAKOW

Critical Early-Stage Decisions in Linked Data Projects for Cultural Heritage: Challenges, Choices, and Guidelines

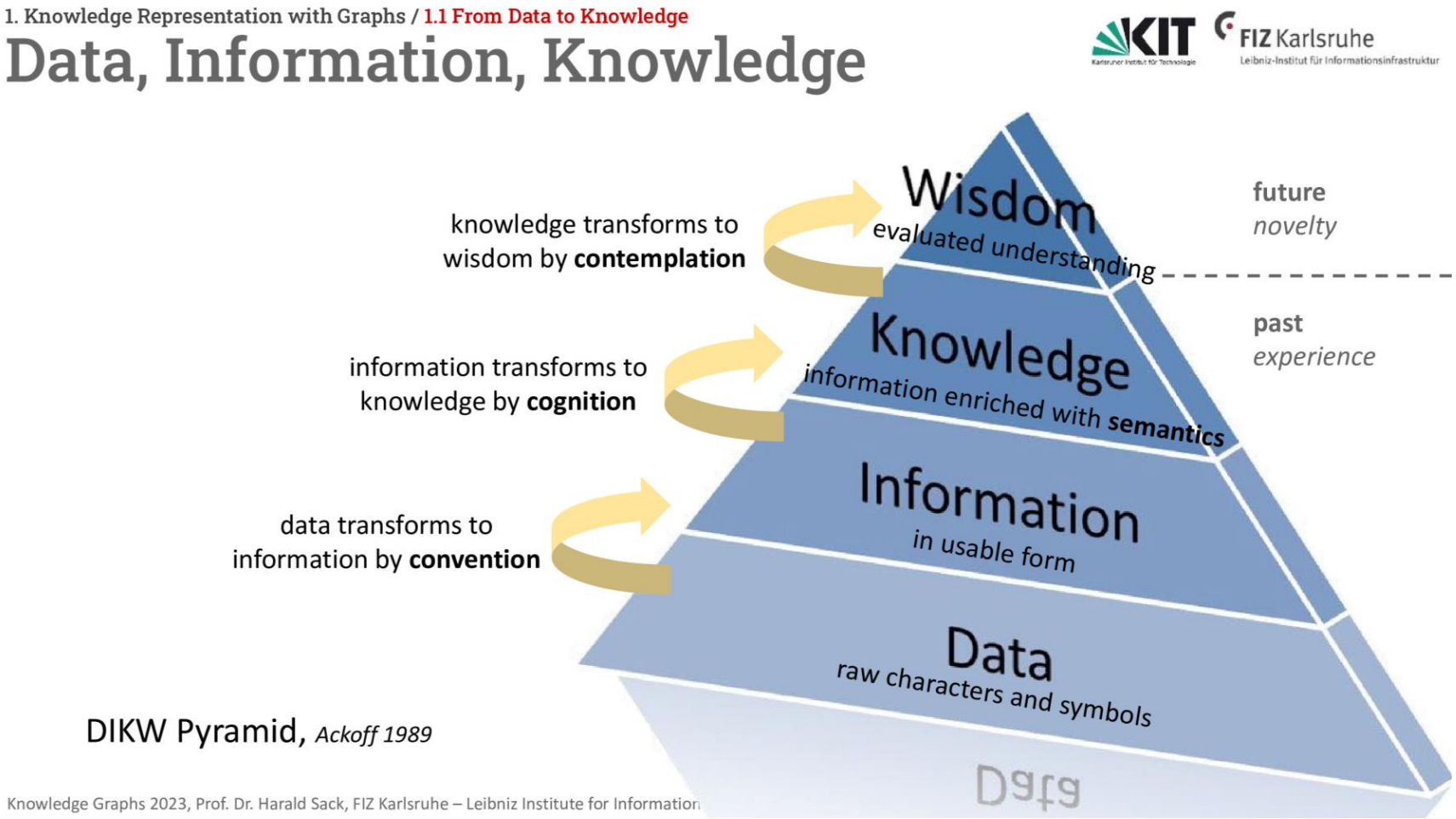
17 October 2024
Luiz do Valle Miranda

Under supervision of: Krzysztoff Kutt and Grzegorz Nalepa
Supported from the CHExRISH project

Outline

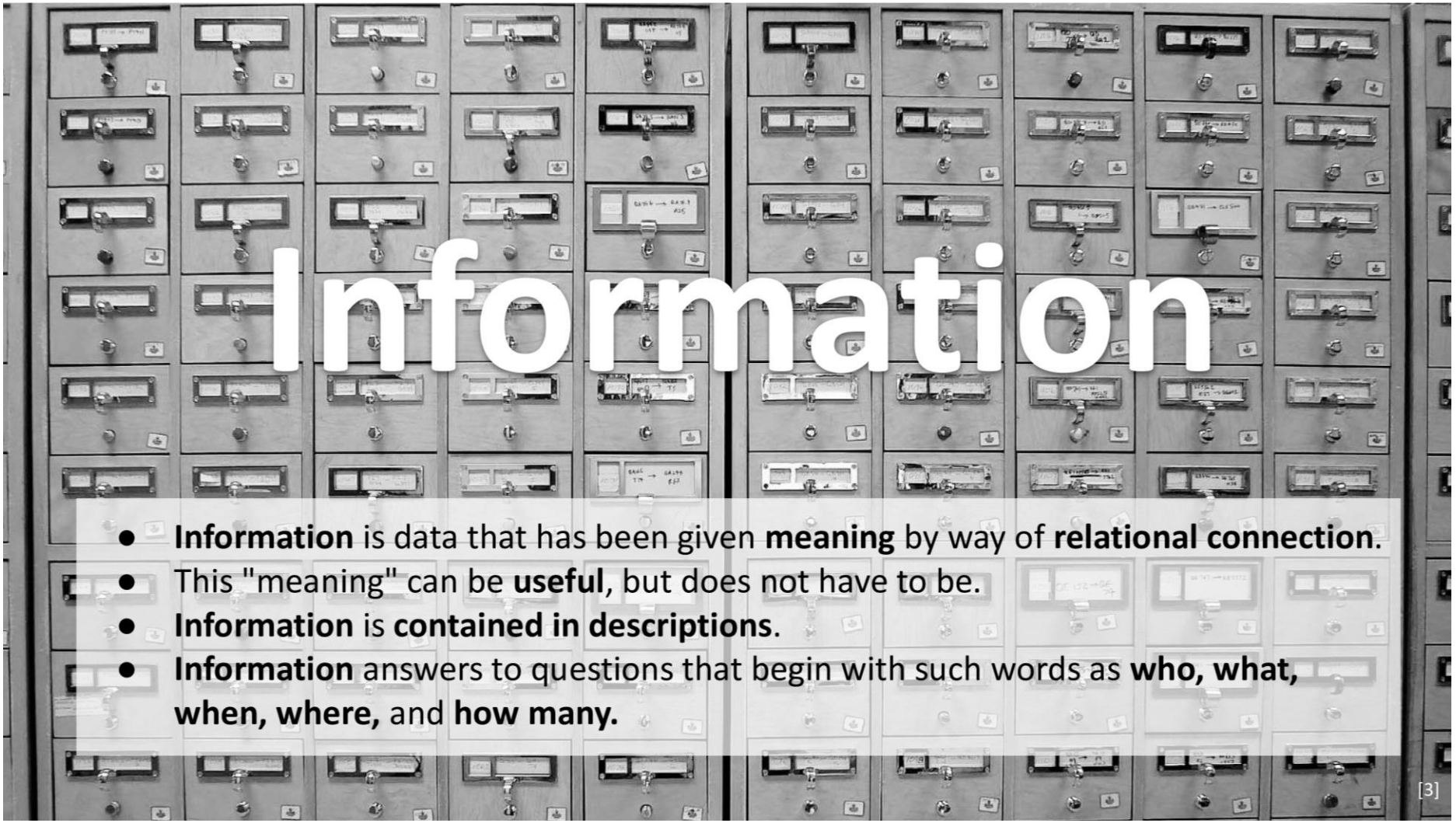
1. Knowledge representation, knowledge graphs, linked data and ontologies
2. Cultural heritage and linked data: affinities, projects and challenges
3. Critical early decisions for LD-based CH projects
4. Planned survey of LD-based CH projects

1. Knowledge representation, knowledge graphs, linked data and ontologies



Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

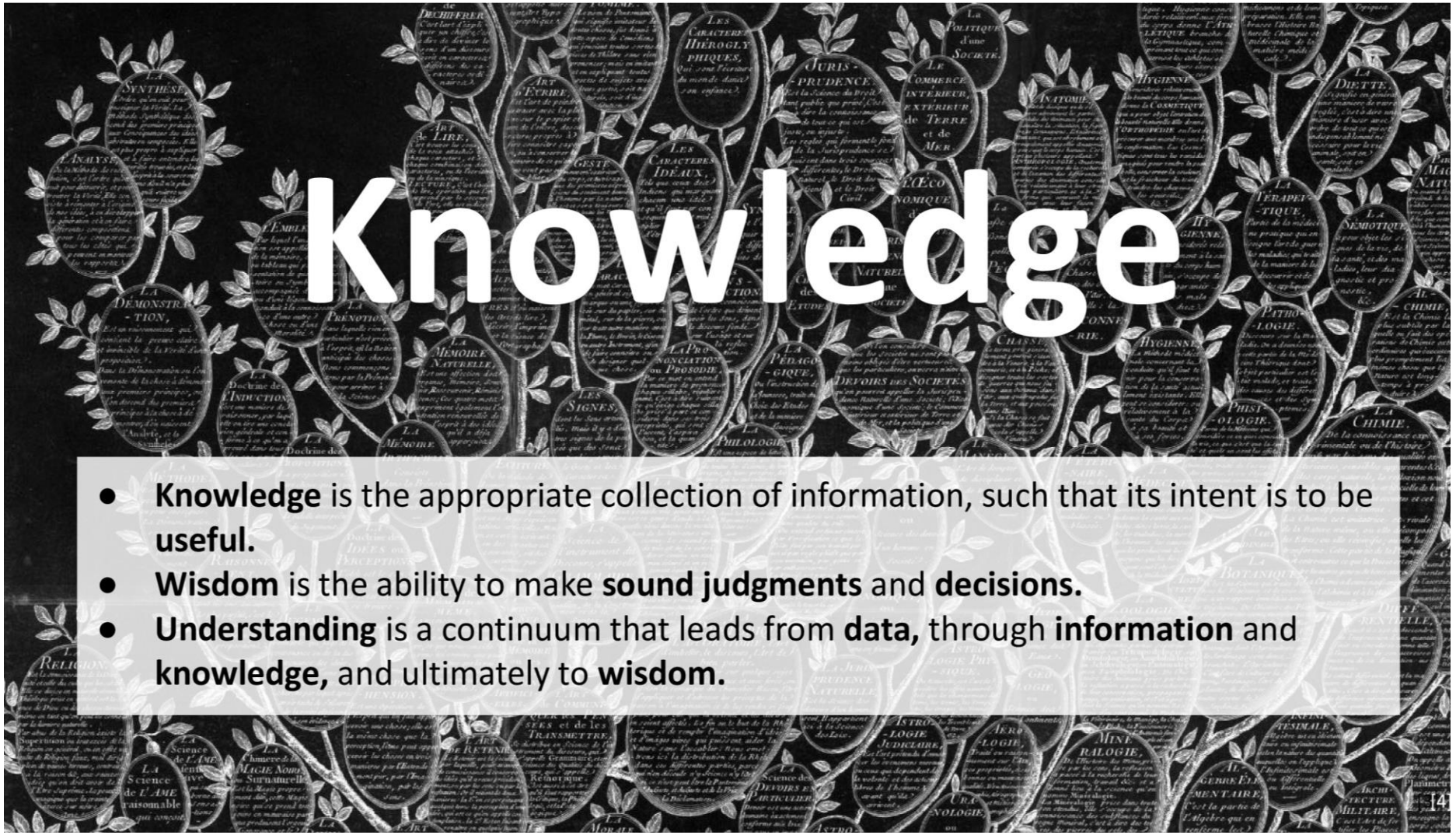


Information

- **Information** is data that has been given **meaning** by way of **relational connection**.
- This "meaning" can be **useful**, but does not have to be.
- **Information** is **contained in descriptions**.
- **Information** answers to questions that begin with such words as **who, what, when, where,** and **how many.**

[3]

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

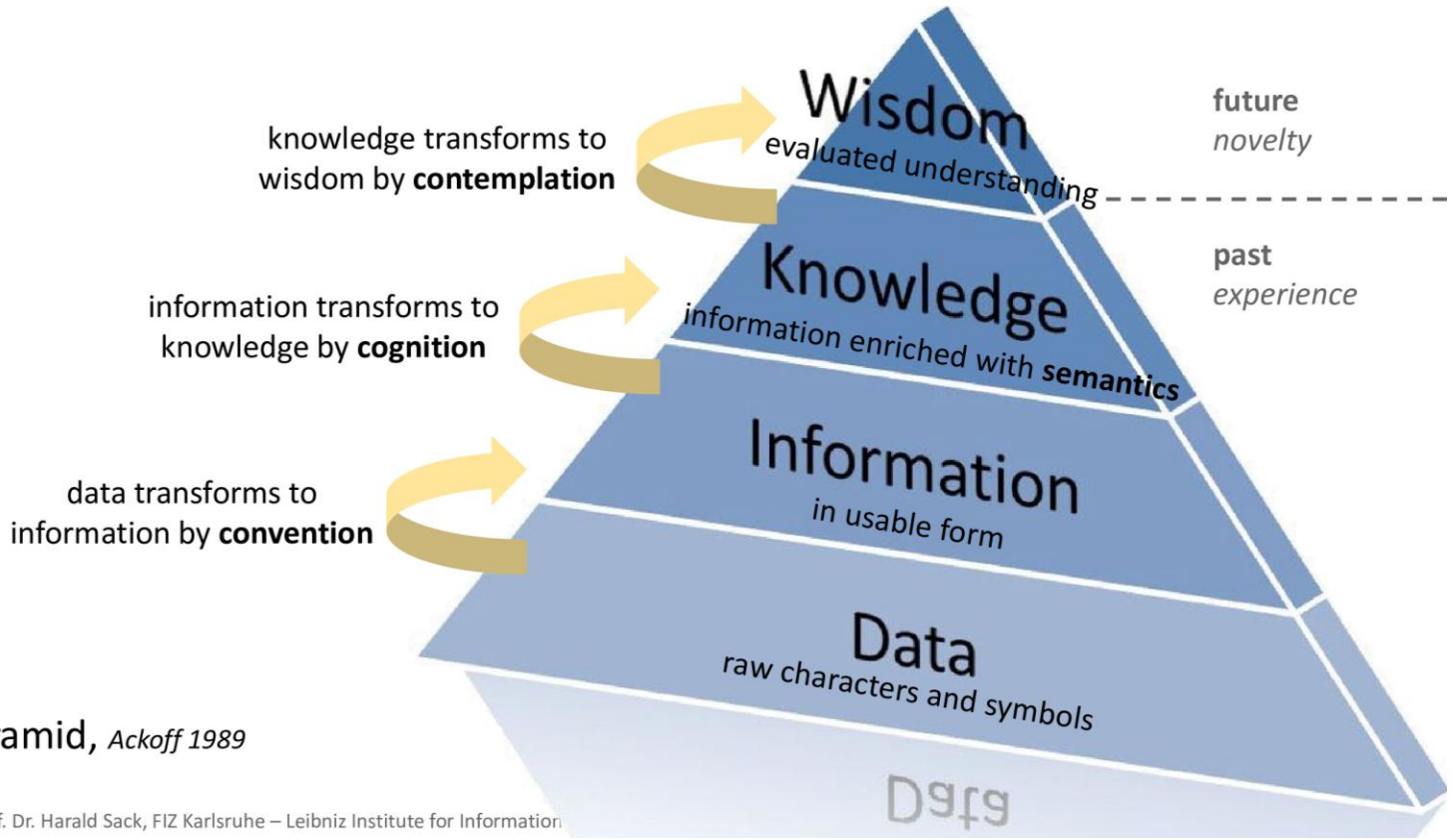


Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

1. Knowledge representation, knowledge graphs, linked data and ontologies

1. Knowledge Representation with Graphs / 1.1 From Data to Knowledge

Data, Information, Knowledge



DIKW Pyramid, *Ackoff 1989*

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

Formal Knowledge Representation

- **Formal Knowledge Representation**
  - is a field of **artificial intelligence (AI),**
  - which (unambiguously) captures the **semantics** of **concepts, properties, relationships,** and **entities**
  - of specific **knowledge domains,** i.e., fields of interest or areas of concern,
  - as **structured data.**
- **Machines (computers)** must be able to **understand** formal knowledge representations.
- To "**understand**" a knowledge representation, the machine must be able to **interpret it correctly**.

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

1. Knowledge representation, knowledge graphs, linked data and ontologies

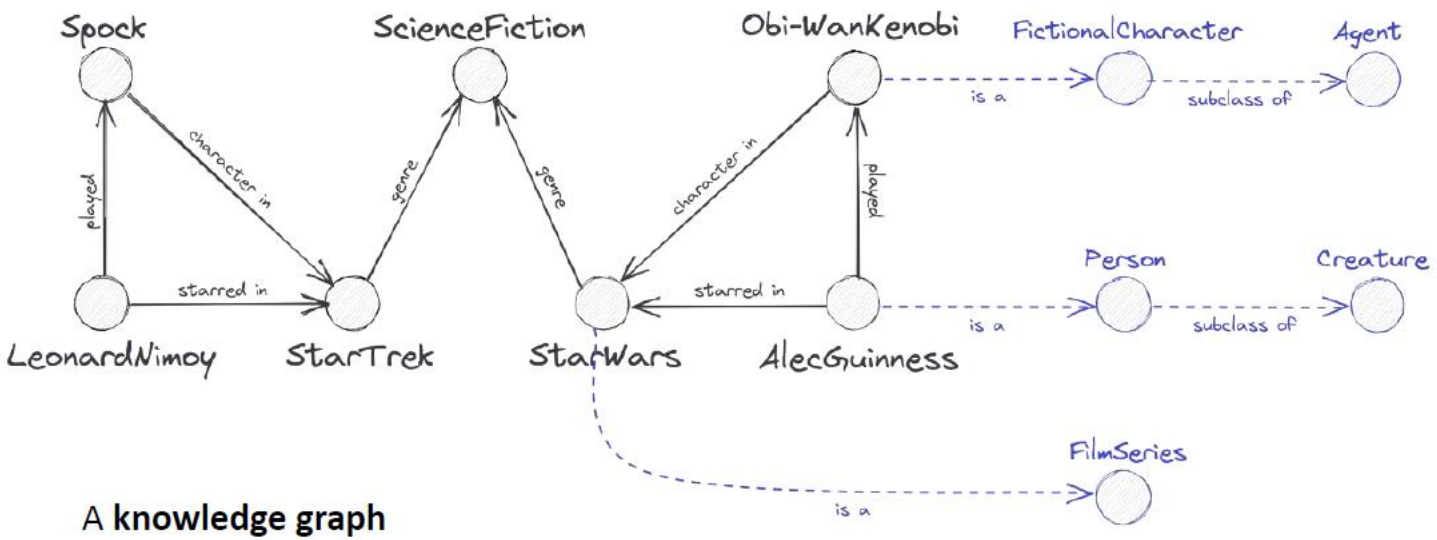
Graphs for meaning representation

- Capture relationships and structures between entities

- Mimics how real-world entities interact and connect

- Can represent both hierarchical and non-hierarchical relationships

Let's Create a Knowledge Graph



A **knowledge graph**

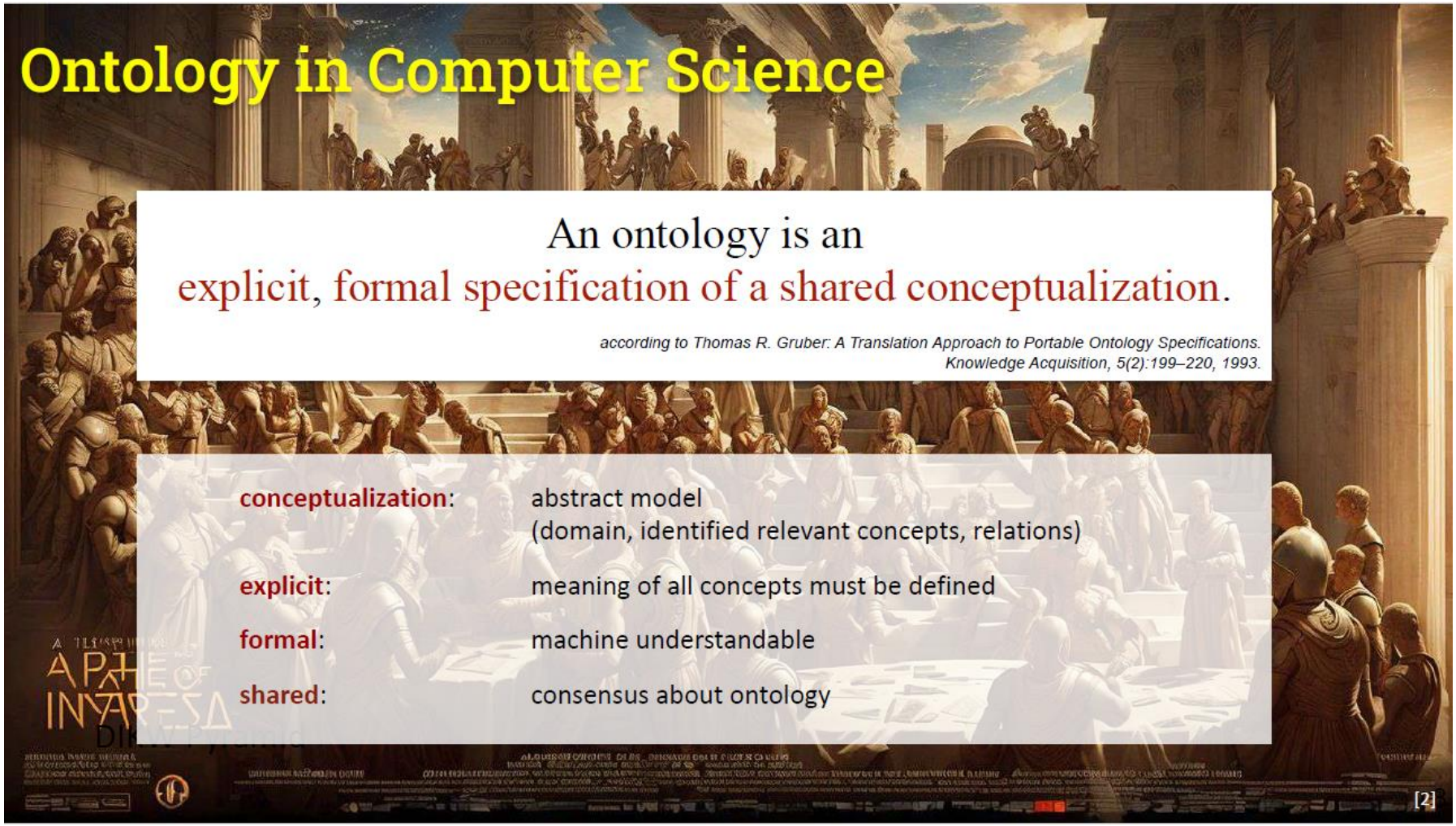
(i) mainly describes real world **entities** and their **interrelations**, **organized in a graph**,
(ii) defines possible **classes** and **relations** of entities in a **schema**,
(iii) allows for **potentially interrelating arbitrary entities** with each other and
(iv) covers **various topical domains**.

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

The Semantic Web – A Web of Data

- The Semantic Web is an **Extension of the traditional Web.**
- The meaning of information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies)**.
- Thereby it will be possible,
  - to **process** the meaning of information automatically,
  - to **relate** and **integrate** heterogeneous data,
  - to **deduce** implicit (not evident) information from existing (evident) information in an automated way.
- The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions**.

Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).



Slide from Knowledge Graphs by prof. Harald Sack (openHPI, 2023).

1. Knowledge representation, knowledge graphs, linked data and ontologies

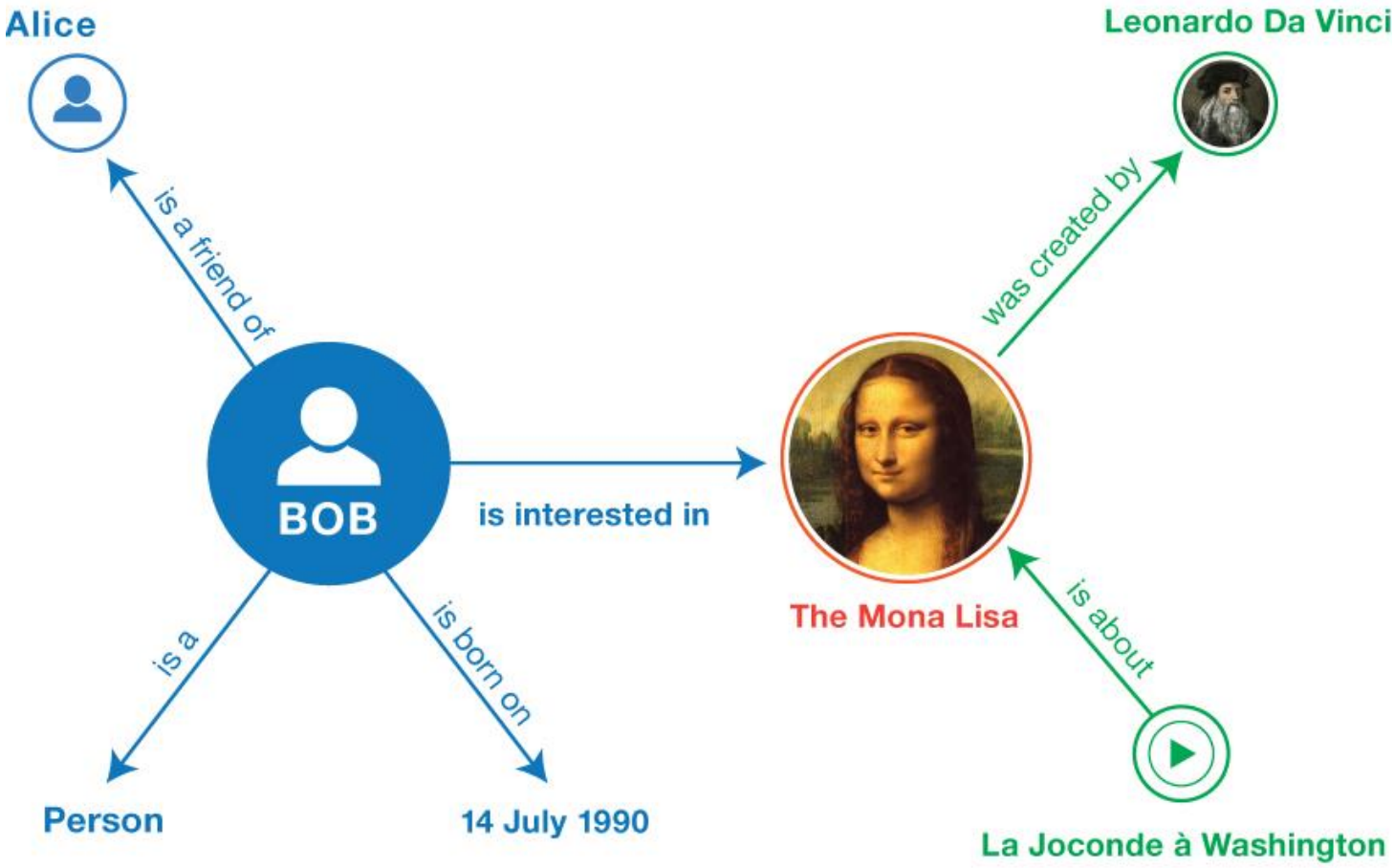


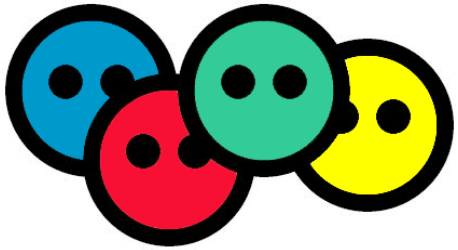
Image from https://wiki.iis.uj.edu.pl/_media/courses:semint:rdf-primer-graph1.jpg

1. Knowledge representation, knowledge graphs, linked data and ontologies

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
         xmlns:dcterms="http://purl.org/dc/terms/"
         xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:schema="http://schema.org/">
   <rdf:Description rdf:about="http://example.org/bob#me">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <schema:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1990-07-
04</schema:birthDate>
      <foaf:knows rdf:resource="http://example.org/alice#me"/>
      <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/Q12418"/>
   </rdf:Description>
   <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
      <dcterms:title>Mona Lisa</dcterms:title>
      <dcterms:creator rdf:resource="http://dbpedia.org/resource/Leonardo_da_Vinci"/>
   </rdf:Description>
   <rdf:Description
rdf:about="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
      <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
   </rdf:Description>
</rdf:RDF>
```

1. Knowledge representation, knowledge graphs, linked data and ontologies

Standards used in Mona Lisa:



Friend of a friend (FOAF)

Dublin Core Terms (dcterms)

Europeana Data Model (EDM)

Wikidata

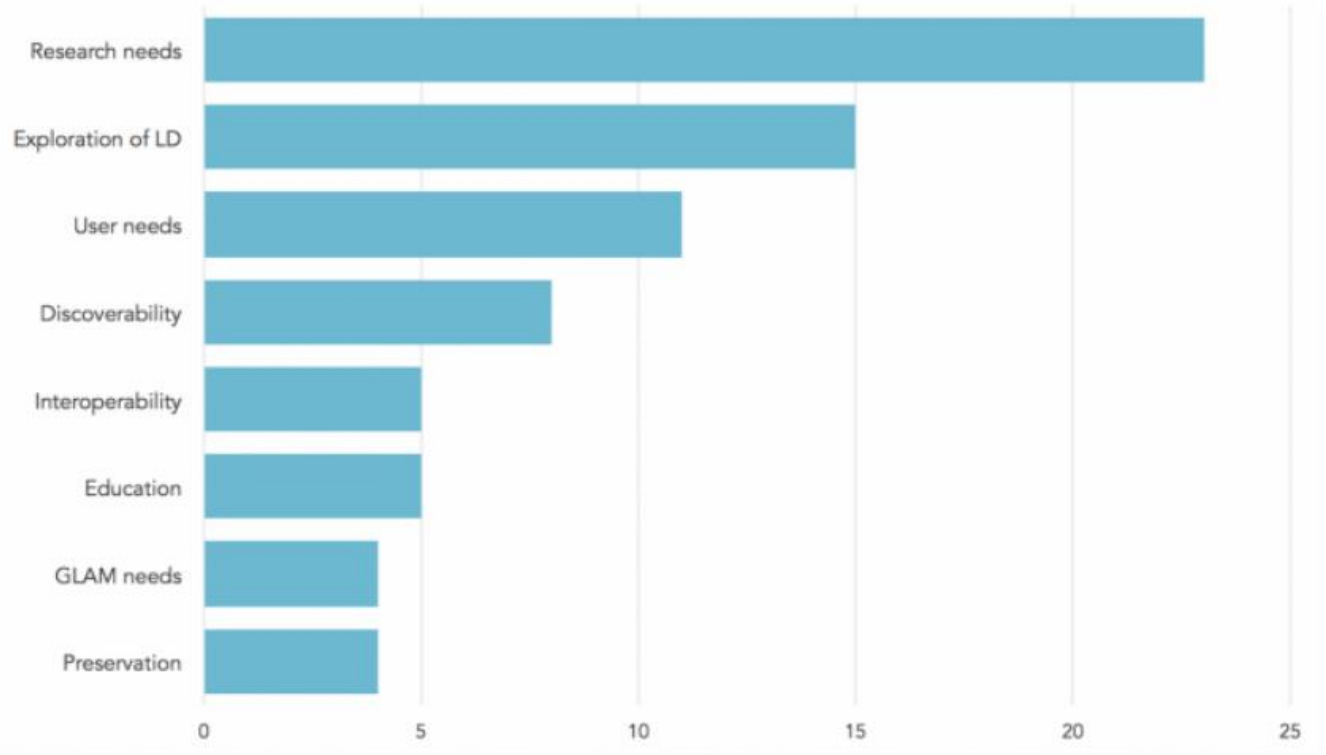
2. Cultural heritage and linked data: affinities, projects and challenges

- Libraries and technology:

- Physical cataloguing vs digital cataloguing

- Linked data

2. Cultural heritage and linked data: affinities, projects and challenges



Motivational factors behind linked data projects in cultural heritage sector

Image from Davis and Heravi (2021)

2. Cultural heritage and linked data: affinities, projects and challenges

1. **Enhanced Information Retrieval**: Semantic search improves the accuracy and relevance of search results by understanding the meaning behind user queries across cultural heritage collections.
2. **Personalized Recommendations**: Linked Data enables context-aware, personalized suggestions for cultural heritage resources based on user preferences and behavior.
3. **Cross-Collection and Interoperable Search:** Linked Data allows users to search across multiple cultural heritage institutions (museums, archives, galleries), connecting resources across different platforms.
4. **Integration with External Knowledge Graphs:** Cultural heritage collections can be enriched by linking to external datasets (e.g., historical records, geospatial data), providing richer context and deeper insights into heritage objects.

2. Cultural heritage and linked data: affinities, projects and challenges

Some significant LD-based CH projects:

europeana
think culture

Europeana

UNIVERSITAS BERGENSIS

The Wittgenstein Archives

universität
innsbruck

Brenner-Archiv

UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

CHExRISH

2. Cultural heritage and linked data: affinities, projects and challenges

Main ontologies for CH

europeana
think culture

Europeana Data Model

CIDOC CRM
CONCEPTUAL REFERENCE MODEL

Cidoc-CRM

ICA

Records in Context – RIC-O

GND

The Integrated Authority File

2. Cultural heritage and linked data: affinities, projects and challenges

Non-CH ontologies often integrated:

Wikidata

Dbpedia

Geonames

SKOS

Simple Knowledge Organization System

2. Cultural heritage and linked data: affinities, projects and challenges

Some challenges for the implementation of LD-based CH projects

1. **Limited Financial and Human Resources**: Smaller institutions like public libraries and local museums often lack the budget and staff to undertake Linked Data projects.
2. **Skill and Educational Gaps**: Lack of adequate training in Linked Data technologies, creating a skills gap that hinders adoption.
3. **Lack of Clear Information and Guidance:** insufficient resources, guidelines, and documentation regarding Linked Data technologies, making it difficult for them to understand how to implement and benefit from these initiatives.

Adapted from Davis and Heravi (2021)

3. Critical early decisions for LD-based CH projects

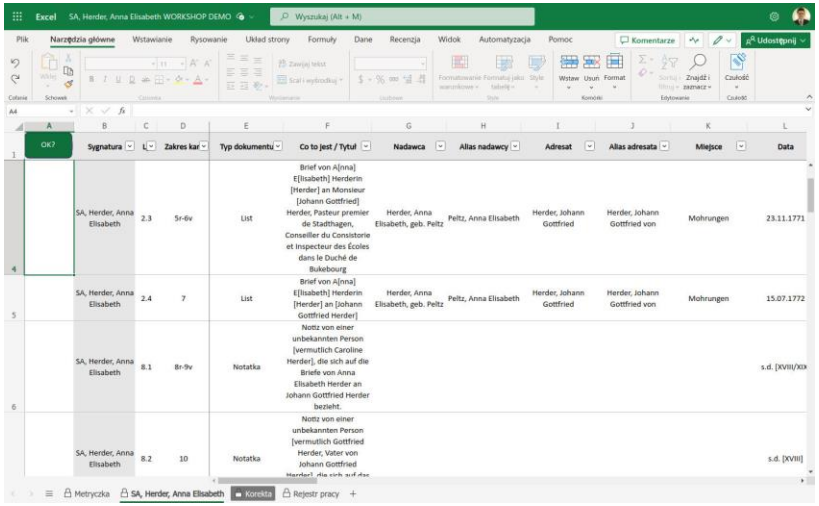
Adapting existing ontology vs creating new ontology

```
<!-- http://www.arcarium.uj.edu.pl/tao/0.3/Agent -->
<owl:Class rdf:about="http://www.arcarium.uj.edu.pl/tao/0.3/Agent">
    <owl:equivalentClass rdf:resource="http://www.europeana.eu/schemas/edm/Agent"/>
    <rdfs:comment xml:lang="en">Person or institution related to documents in various ways (eg. creator, sender, recipient, issuer, obtainer)
                                [record that appears in SEC]. Abstract class</rdfs:comment>
    <skos:scopeNote xml:lang="en">This is a TAO class</skos:scopeNote>
</owl:Class>

<!-- http://www.arcarium.uj.edu.pl/tao/0.3/CollectionEntity -->
<owl:Class rdf:about="http://www.arcarium.uj.edu.pl/tao/0.3/CollectionEntity">
    <owl:equivalentClass rdf:resource="http://www.europeana.eu/schemas/edm/ProvidedCHO"/>
    <rdfs:label xml:lang="en">Collection Entity</rdfs:label>
    <skos:definition xml:lang="en">Class to represent different kinds of objects that can be singled out in the collection</skos:definition>
    <skos:scopeNote xml:lang="en">This is a TAO class</skos:scopeNote>
</owl:Class>
```

3. Critical early decisions for LD-based CH projects

How to collect data from different stakeholders?



omeka s

3. Critical early decisions for LD-based CH projects

How to query linked data: local copy vs online querying?

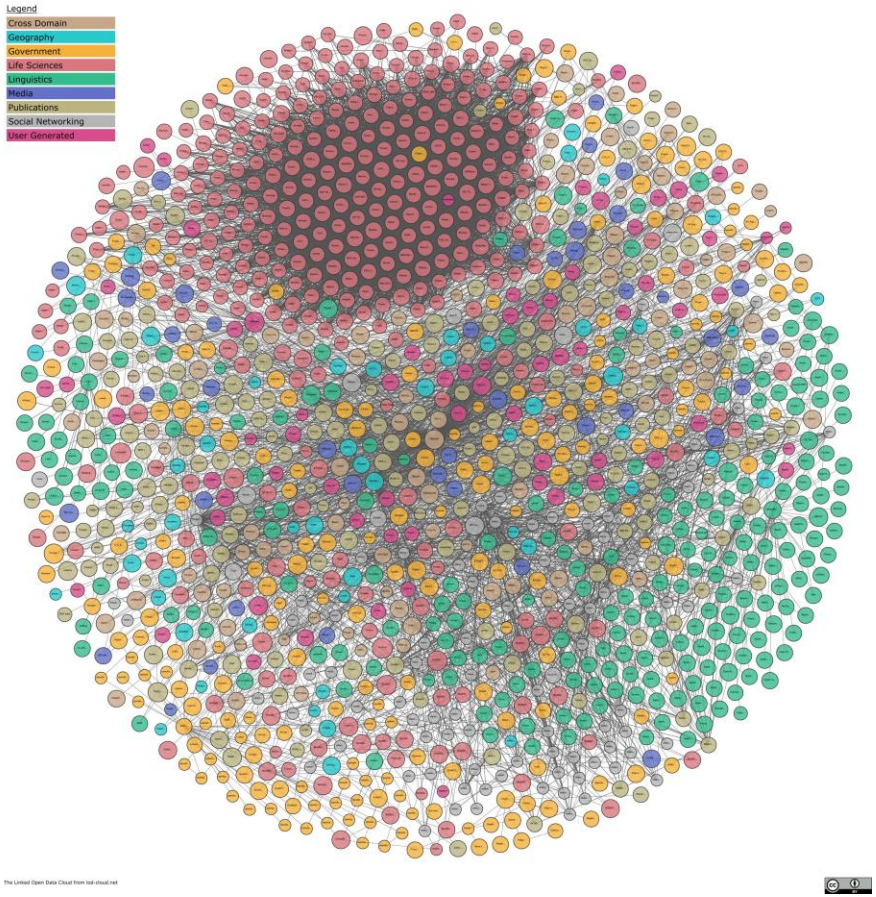


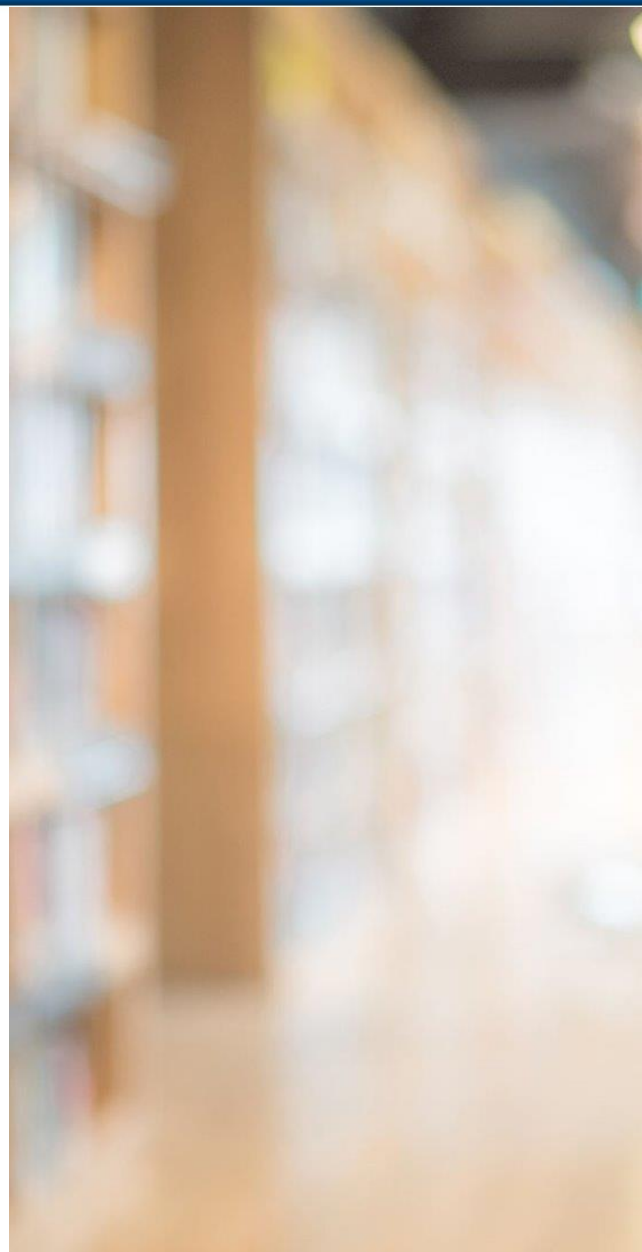
Image from https://lod-cloud.net/

3. Critical early decisions for LD-based CH projects

Strategy to ensure long-term:

- Preservation
- Sustainability
- Access

to digital artifacts



3. Critical early decisions for LD-based CH projects

Choice between open vs closed data models

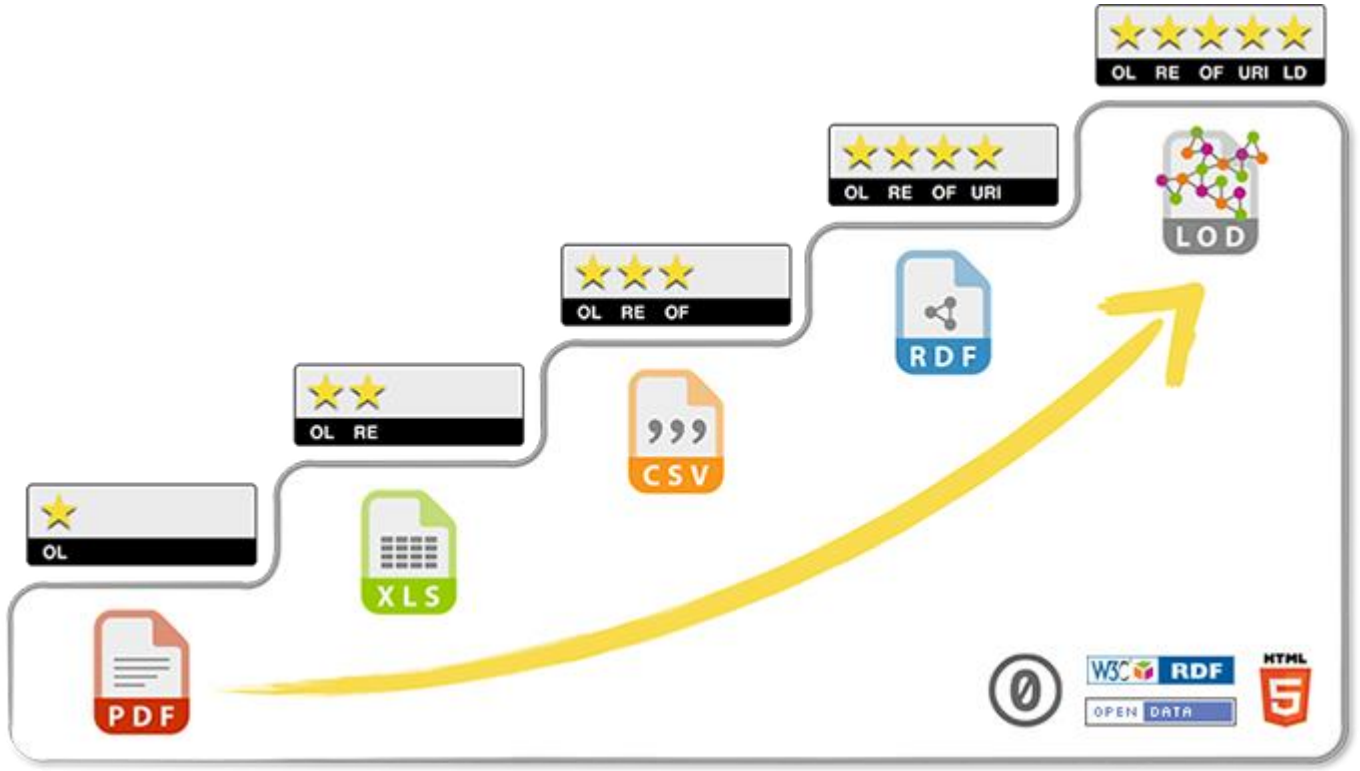
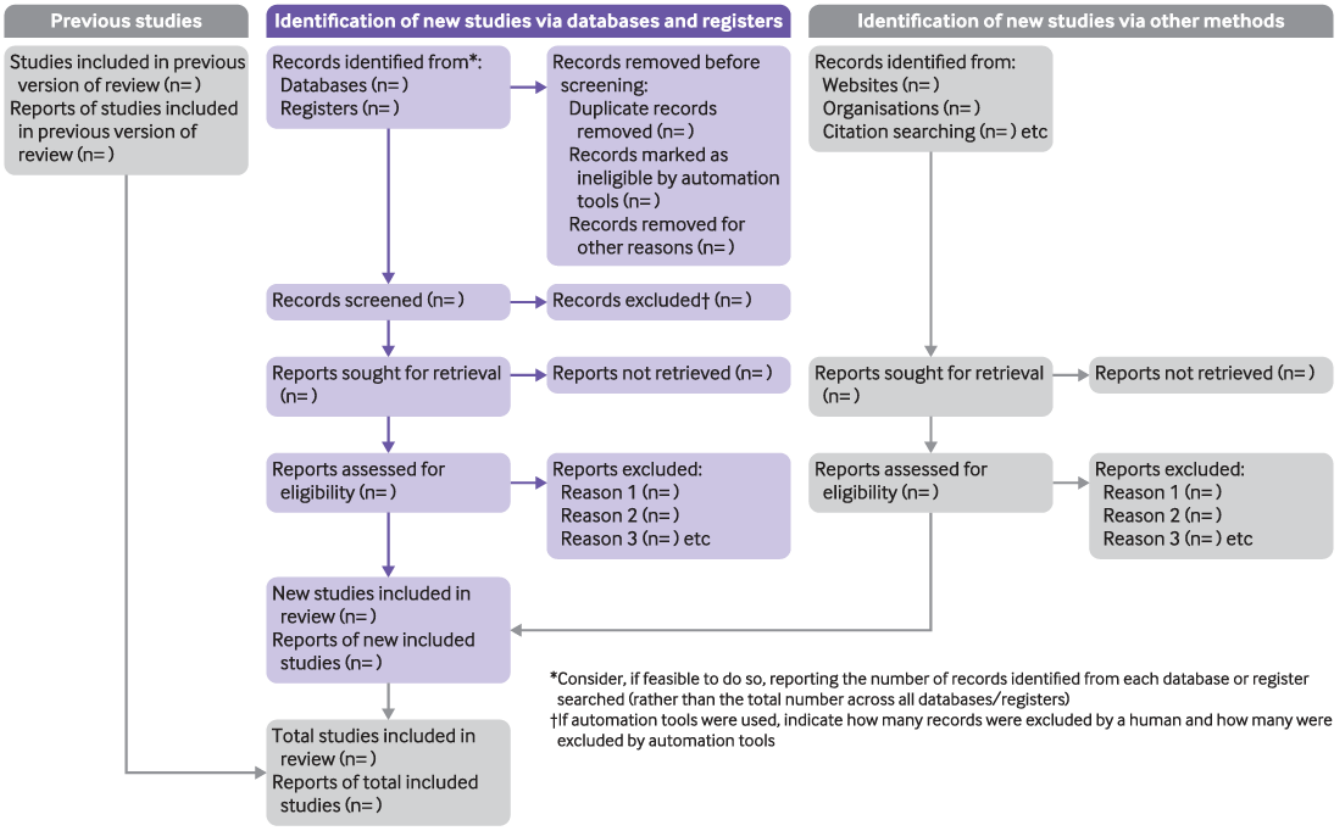


Image from Hitz-Gamper et al. 2019

4. Planned survey of LD-based CH projects



PRISMA 2020 flow diagram template for systematic reviews.

Image from Page et al 2021

4. Planned survey of LD-based CH projects

Research questions:

R1: In what cases would new ontologies be created or existing ones be adapted?

R2: How to choose the best tools for data collection?

R3: How to ensure the long-term preservation and sustainability of digital artifacts within CH systems?

R4: How to query linked data: local copy vs online querying?

R5: What are the stakes in choosing open vs closed data model

4. Planned survey of LD-based CH projects

Keywords for literature search:

Group 1: Cultural Heritage Projects

Group 2: Linked data

Group 3: ?

Thank you for your attention

Comments?

Bibliography

Davis, E., & Heravi, B. (2021). Linked Data and cultural heritage: A systematic review of participation, collaboration, and motivation. Journal on Computing and Cultural Heritage, 14(2), Article 21. https://doi.org/10.1145/3429458

Hitz-Gamper, B., Neumann, O., & Stürmer, M. (2019). Balancing control, usability and visibility of linked open government data to create public value. International Journal of Public Sector Management, 32. https://doi.org/10.1108/IJPSM-02-2018-0062

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Systematic Reviews, 10(1). https://doi.org/10.1186/s13643-021-01626-4