

Knowledge Graphs for Digital Libraries

A (more) holistic access to research artifacts

Presenter | Fidan Limani

Venue | LIBER 2021

Date | June 24, 2021

Outline

Scholarly communication changes

The Knowledge Graph Approach

Use Cases for DLs

Conclusion & Future work

Scholarly communication changes

- Technical and cultural changes
- Library domain opportunities
- Meanwhile at the ZBW...

Technical and cultural changes

IT-empowered Research Lifecycle

- Increased (technical) capabilities, at all levels (personal, organizational, community)
- Infrastructures of different **scopes** (national and international) and **models** (free, open, commercial), etc.

More, faster, better

- Do more in the course of a research lifecycle
- Generate more (both volume and variety), store more, process more, etc.

Getting the whole research picture

- Bridge the isolated artifacts collections
- Bring together artifacts from the same research/different types of deliverables/artifacts from a domain.



www.shutterstock.com · 1047792202

Figure 1. Siloed research artifacts.
Adapted from [cigdem](#) ([link](#))

Library Domain Opportunities

Requests for new research artifacts

- It used to be only (mainly?) research articles...
- Emerging artifacts: Sci. blogs & wikis, research data, links between artifacts, social events, etc.
- Requests coming from different stakeholders: Authors, researchers, library users, etc.

Relevant initiatives at the ZBW

- Literature artifacts: EconBiz, Econstor, Journal Data Archive
- Research data:
 - GeRDI: Long tail research data from different disciplines,
 - KonsortSWD: Research data from social sciences,
 - Other non-structured, or data not aimed for research, etc., are also being targeted.

Should we bother to engage?

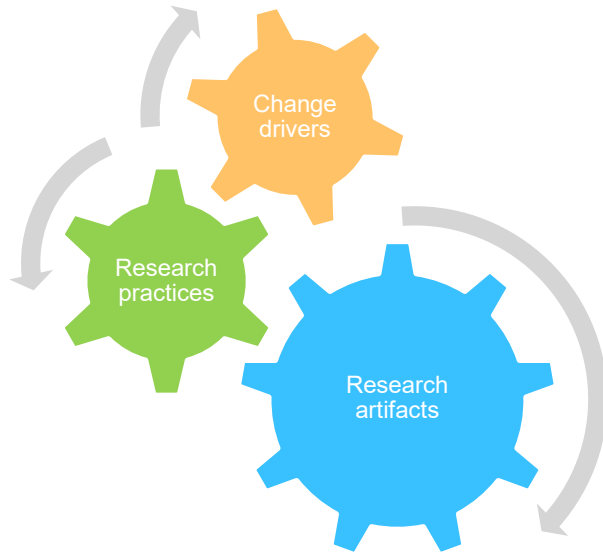


Figure 2. Research Artifacts:
Increasing in importance

Inter-play of incentives

- Raw data captured and published;
- Its documentation via a **data paper**, for e.g.;
- Data generation/gathering credited (in addition to article publication)
- An incentive to generate/collect and publish more raw data

All research artifacts are welcome!

- Research data
- Configuration scripts
- Scientific notebooks, etc.

Meanwhile at the ZBW ...

Artifacts status quo

- Ever-increasing types on radar (Ref[1, 2, 3])
- Different management practices, such as metadata standards
- Structuring
- Terminology
- Even with the same domain

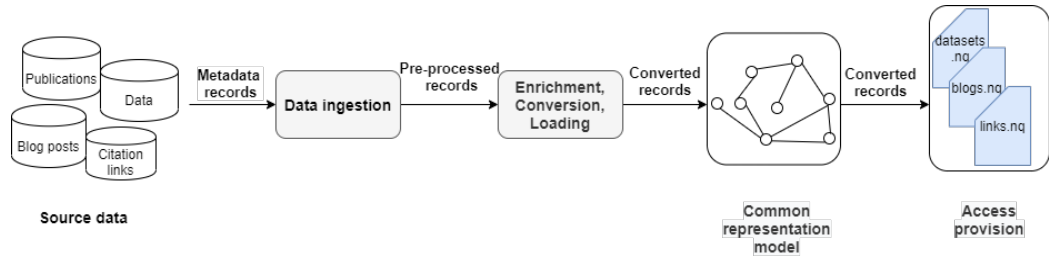


Figure 3. Artifacts conversion workflow

Initial take & findings

- Bring artifacts on a common representational model
- Aligning them terminology-wise remains challenging
- Mixed results across different artifacts

Abstracting away...

- Scholarly communication keeps changing
- So will the requests for new artifacts
- Having a common approach is important



[4, 5]

The Knowledge Graph Approach

- KGs: A Definition
- Why KGs?
- Meanwhile at the ZBW...

KGs: A Definition

What is a KG?

- Not a single definition accepted across the communities (Good? OK?)
- Ehrlinger and Wöß [Ref 4] collect several definition categories:
 - “Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities.” [Ref 5]; or
 - “... a knowledge graph describes objects of interest and connections between them” [Ref 6]

The KG Emergence

- Open vs. Closed: Wikidata, Dbpedia vs. (Google, Microsoft, Facebook, et al.) KG, etc.
- Enterprise KGs: Tailoring to the needs of an enterprise
- General vs Domain specific
- Academic KGs: Microsoft Academic Graph, The Open Research KG, Information Space Graph, etc.

Why KGs?

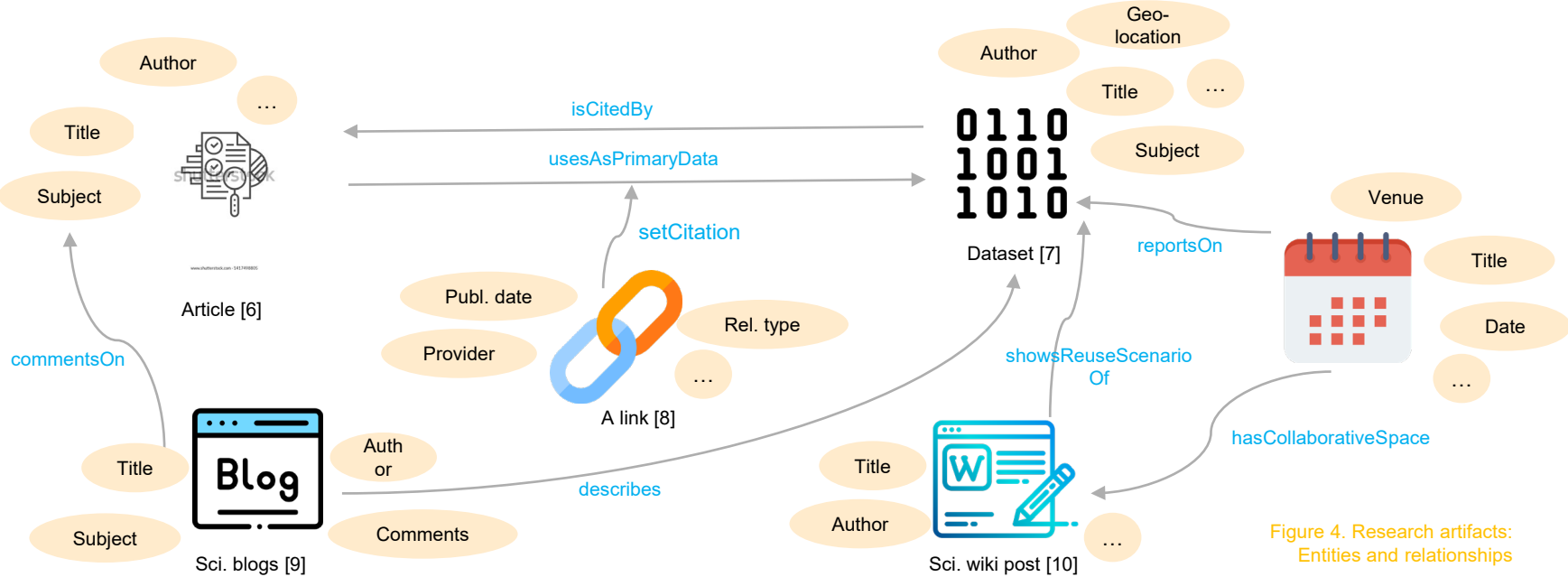


Figure 4. Research artifacts: Entities and relationships

Why KGs? (Cont.)

Resource of interest

- Research artifacts of different types
- Continuous emergence of new ones

Benefits from adopting KGs

- Integration as a key feature
- Support extensibility
- Accepts heterogeneity

KGs: A good fit

- Represent entities and relationships
- Support types for entities and relationships
- Combine with existing graph collections, etc.

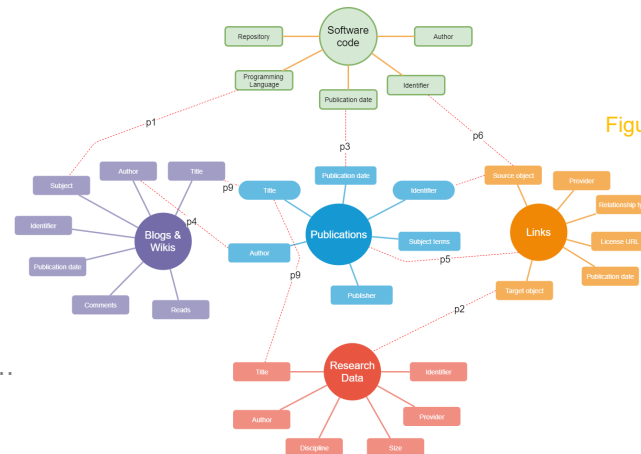


Figure 5. KGs are quite suitable to represent research artifacts

Meanwhile at the ZBW...

(Open Access) Publications

- EconStor: <https://www.econstor.eu/>
- **108 K** + publications

Research data

- Project GeRDI: <https://www.gerdi-project.eu/>;
> **1.1 M** metadata records;
- Journal Data Archive (<https://journaldata.zbw.eu/>);
< **200** metadata records;

Links between scholarly artifacts

- Schoexplorer (<http://schoexplorer.openaire.eu/>)
> **126 M** citation links, both literature-to-dataset
and dataset-to-dataset
- **17 providers**

Blog posts

- VoxEU (<https://voxeu.org/>)
- **8.5 K** blog posts

The KG components

Data source

- The original source collections

Data ingestion

- Ingest/harvest the source metadata

Structuring layer

- STW application

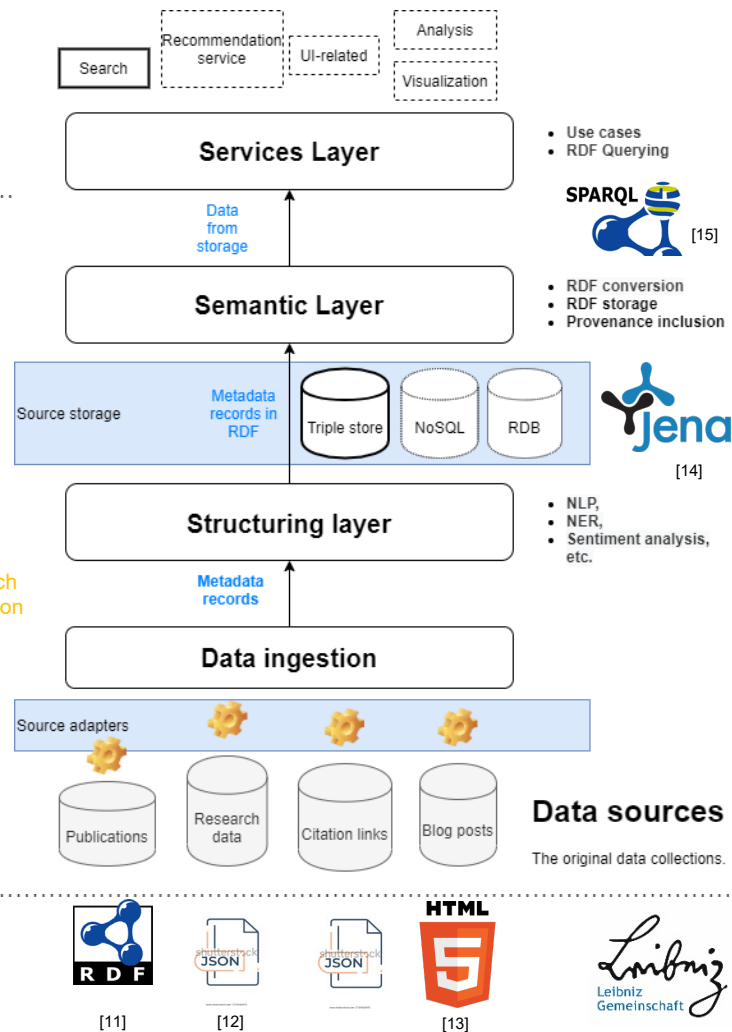
Semantic layer

- Ontology/Vocabulary selection
- Ontological description of the resources

Services layer

- Services to support DL use cases

Figure 6. A KG instance for 4 research artifact types under exploration



Use Cases for DLs

- Current focus
- Use cases for individual and cross artifact types

Current focus

At the ZBW, we are considering the KG adoption

- Exploring the technology solutions for the KG
- Identifying activities that operate on the KG collection (enrichment, linking, etc.)
- The focus: Enterprise-type KG for the artifacts the ZBW curates

In the process, it is

- Important to ID use cases
- Necessary to include all the relevant stakeholders

Do you have a use case? 😊

Use Cases

Individual artifacts

- Blog posts & Wikis
- Publications
- Research Data
- Citation Links
- Events

Multiple artifacts

- Resources linked by the same publisher, publication date, etc.
- Top datasets cited by publications and blog posts
- Datasets/publications on a topic that a publication/dataset addresses.
- The most cited authors (publications, datasets) in a domain
- The most cited events (publications and datasets) for a given time period, domain, etc.

Note All this depends on the available metadata!

Conclusion & Future Work

- KGs for a (more) holistic approach to research access
- Future work

Conclusion

KGs as means for a (more) holistic approach to research

- Publication & Dissemination
- A “one-stop shop” approach to accessing research deliverables
- Tailored to the domain of economics and social sciences
- Unlocking interesting use cases

Collaborative KGs

- Different scopes, (meta)data granularities, purposes, etc., captured with different KGs
- A bottom up approach to a more complete Web of Data
 - The collaboration b/w different KGs makes this a viable approach

Future work

KG components

- Data source: Include new sources
- Data ingestion: Support new (harvesting) protocols and cases
- Structuring layer: Do more with the available metadata
- Semantic layer: Ontologies that represent a more “packaged” way of publishing research artifacts + Validation

Other aspects

- Establish links to other KG: A “Web of Data” based on multiple (enterprise, domain-specific, etc.) KGs;
- Enrichment
- Link up
- Update and Maintenance, etc.

Questions / Comments?

Thank you for your attention!

E-mail: f.limani@zbw.eu

References

- [Ref 1]** Limani, F., Latif, A., & Tochtermann, K. (2019, October). Scholarly Resources Structuring: Use Cases for Digital Libraries. In Research Conference on Metadata and Semantics Research (pp. 248-259). Springer, Cham.
- [Ref 2]** Limani, Fidan, Atif Latif, and Klaus Tochtermann. "Linked publications and research data: use cases for digital libraries." International Conference on Theory and Practice of Digital Libraries. Springer, Cham, 2018.
- [Ref 3]** Limani, F., Latif, A., & Tochtermann, K. (2019, October). Links between research artifacts: Use cases for Digital Libraries. In International Journal of Metadata, Semantics and Ontologies (Accepted).
- [Ref 4]** Ehrlinger, L. and Wöß, W., 2016. Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48, pp.1-4.
- [Ref 5]** M. Kroetsch and G. Weikum. Journal of Web Semantics: Special Issue on Knowledge Graphs.
<http://www.websemanticsjournal.org/index.php/ps/announcement/view/19> [August, 2016].
- [Ref 6]** Natasha F. Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. 2019. Industry-scale Knowledge Graphs: Lessons and Challenges. *ACM Queue* 17, 2 (2019), 20.