LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

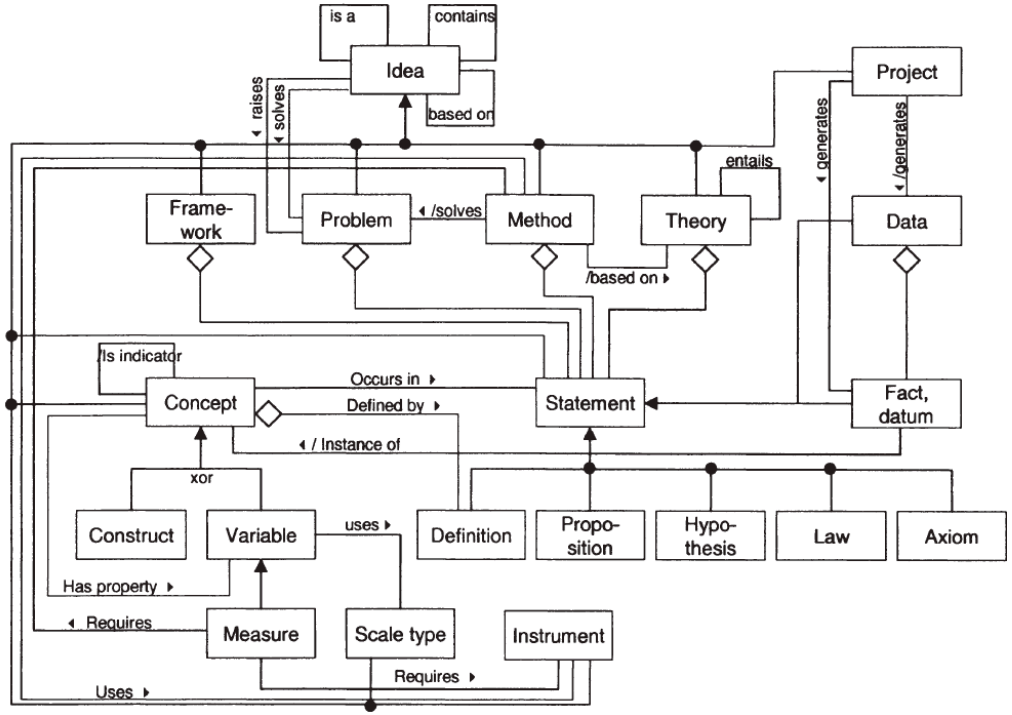
Seven challenges of machine actionable scholarly knowledge

Markus Stocker

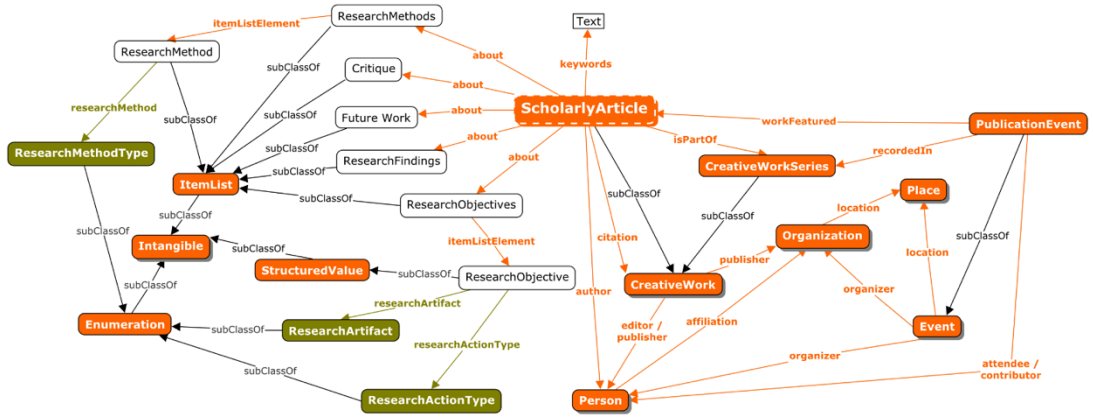
Leibniz Gemeinschaft

Knowledge Model

- What is the structure of scientific knowledge?
- What is universal, if anything, and what is domain specific?
- How do we get to a model: Top down, buttom up, a mixture?



Hars A. (2003) Structure of scientific knowledge. In: From Publishing to Knowledge Networks. Springer, Berlin, Heidelberg
https://doi.org/10.1007/978-3-540-24737-1_3



Vera G. Meister (2017). Towards a Knowledge Graph for a Research Group with Focus on Qualitative Analysis of Scholarly Papers. In Proceedings of the First Workshop on Enabling Open Semantic Science. http://ceur-ws.org/Vol-1931/paper-10.pdf

… among many others

Tractable Granularity

- Some of these knowledge models are relatively complex: 20+ classes and 40+ relations
- The ORKG uses a very simple knowledge model: A **ResearchContribution** is an entity that relates the **ResearchProblem** addressed by the contribution with the **ResearchMethod** and (at least one) **ResearchResult**
- The granularity of these knowledge models vary considerably
- At what granularity does it become intractable/impractical to instantiate knowledge models?
- Does this differ for different content acquisition approaches?

Content Acquisition

	Content acquisition modes ...		
Content acquisition at the time of ...	**Crowdsourced**	**Automatically**	**Infrastructurally (*)**
Submission/Publication	●	●	
Writing	●	●	●
Reading	●	●	
Knowledge production			●
Legacy		●	

(*) In terms of advanced research infrastructure

Multimodality

- A multi-modal approach to content acquisition is necessary
- Each mode has its place, pros and cons
- More importantly, we need multiple modes to instantiate granular knowledge models
- Integrating the various modes, as they are executed during the research lifecycle, is an engineering challenge

Crowdsourcing Content

- Crowdsourcing has been used in numerous contexts
- Online Encyclopedia and Maps are good examples
- In these contexts, the ratio between content producers and content consumers is extremely low
- “1-9-90 rule”: 90% only view content, 9% edit content, 1% create new content
- The scholarly context may be different
- Scholarly knowledge is so specialized that we probably rely on more content producers
- **More importantly, are non-authors legally allowed to create content published by others?**

Content Storage

- Knowledge Graphs? And what are these anyway? Triple Stores, Labeled Property Graph, …?
- Why are we employing one technology over the other?
- The ORKG is already employing a mixture of storage technologies
- Including plain old relational databases
- Will we realize that operational infrastructure is after all better built on technology that has withstood the test of time?

Content Linking

- To what third-party content can we link scholarly knowledge?
- Some ORKG use cases make use of third-party terminologies
- The RDF Data Cube Vocabulary is currently the canonical vocabulary to represent datasets
- Should we represent datasets or link to them? What are the implications?
- Should a software implementing a method be described or simply linked?
- Is a video “related to” a research contribution or a manifestation of a research contribution?