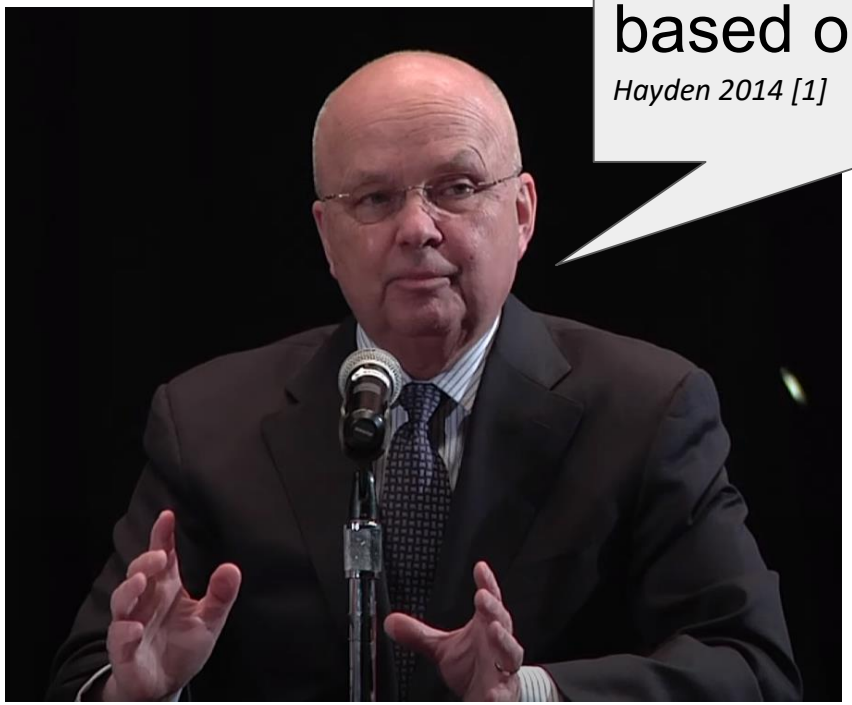
Semantic MediaWiki for community-wide Research Data Management

Fabian Hoppe, Tabea Tietz and Harald Sack

Int. Workshop on FAIR Data in Plasma Science, 16.05.2022

Why is metadata important?



General Michael Hayden,
former director of the CIA and NSA

Metadata provides enough knowledge for the NSA to decide whom to kill.

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

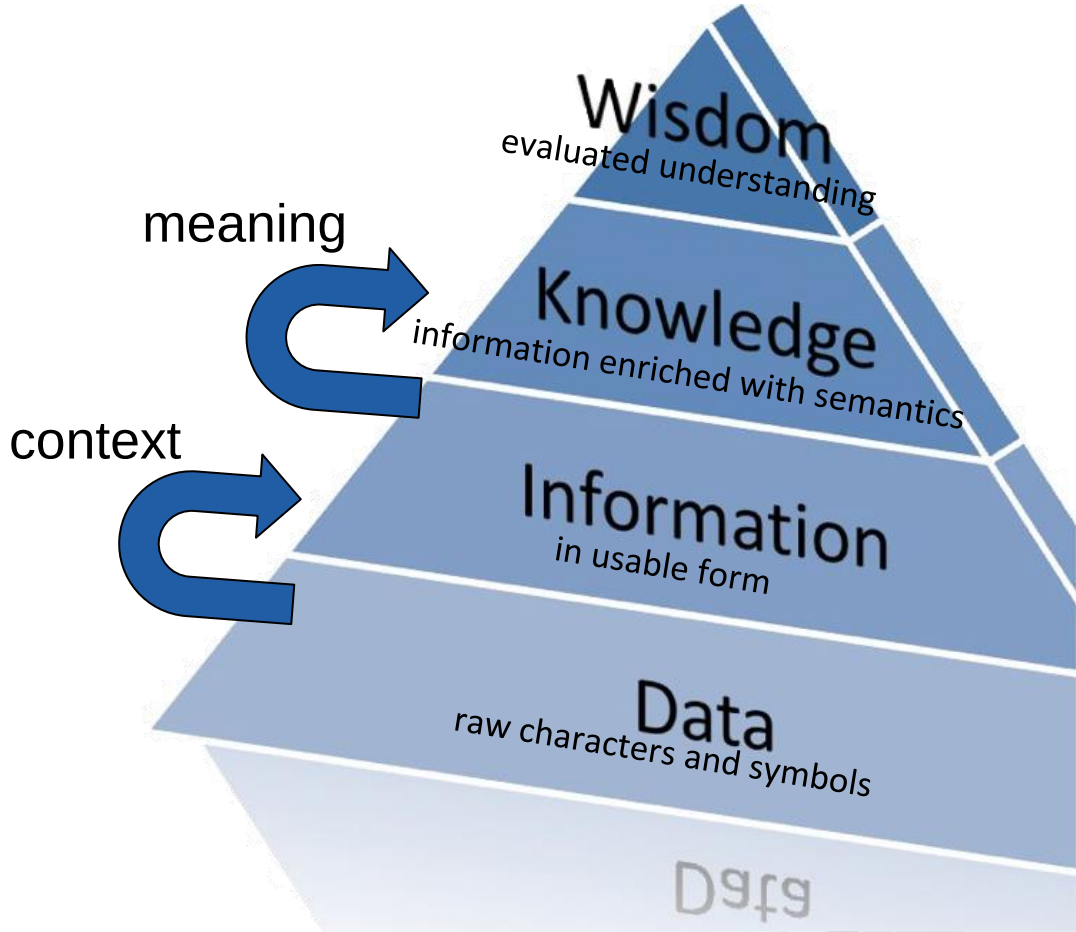
Why is **my** metadata important?

Get knowledge to make decisions

- What measurement device was used?
- Can I reproduce this experiment?
- Which experiments used kINPen® IND?

Knowledge

- Collection of information with intent to be useful
- Information
  - data that has been given meaning by way of relational connection



DIKW Pyramid, *Ackoff 1989 [2]*

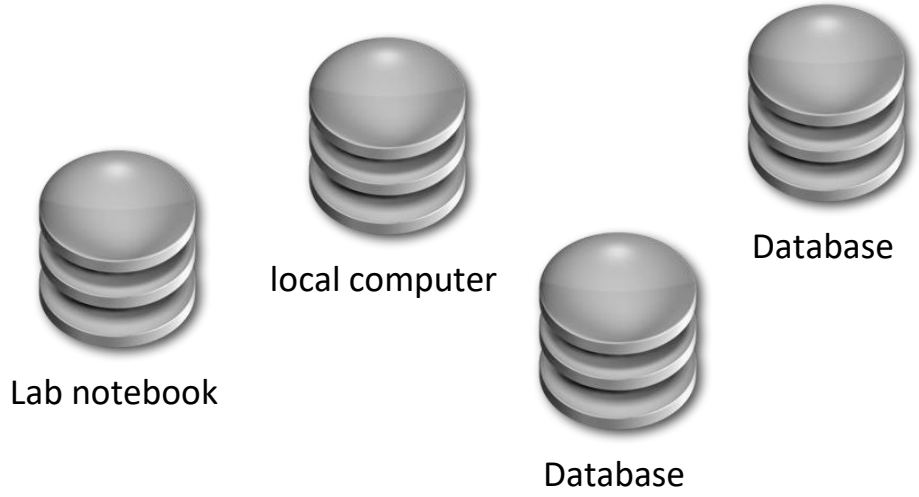
Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

How do we get there?

Data is locked up in small data islands

- different formats, schemas
- no links between data

→ even saved metadata is hard to access



Make (meta)data **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

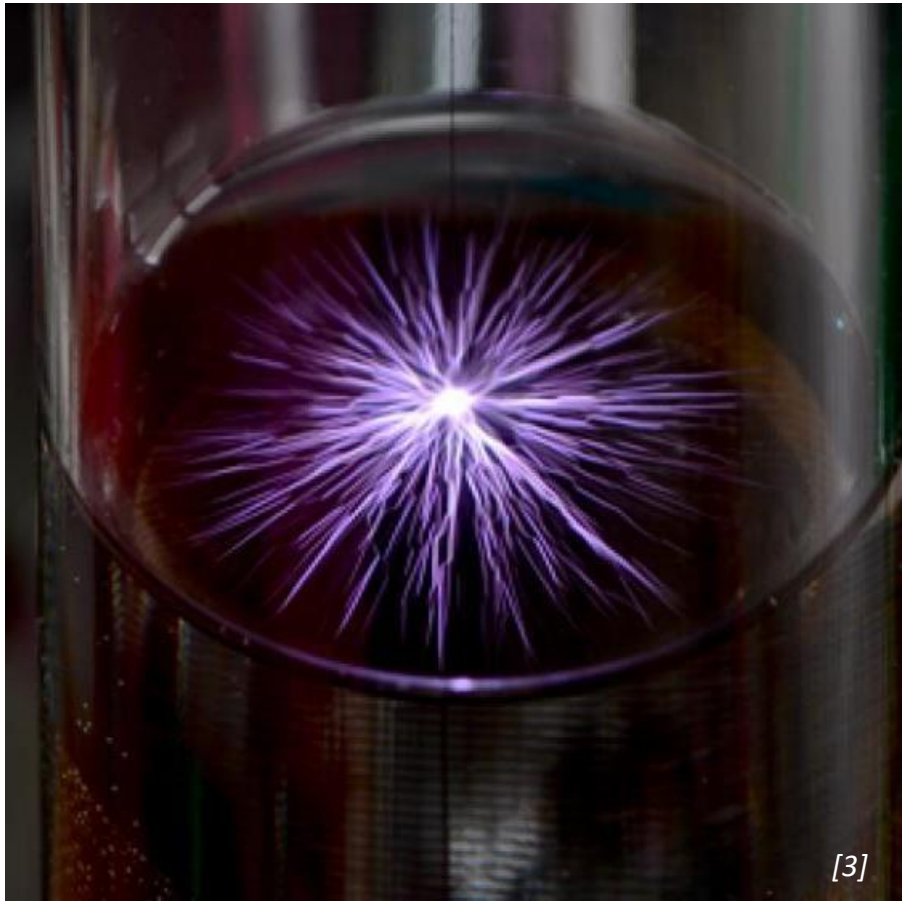
- Apply linked data technology
- Make data machine-accessible, machine-understandable

→**formal knowledge representation**

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Overview

- Ontology Development
  - What is an ontology?
  - Advantages
  - Development process
- Community driven Development
  - Why is it needed for RDM?
  - Requirements
- Semantic MediaWiki
  - Workflow
  - Interface

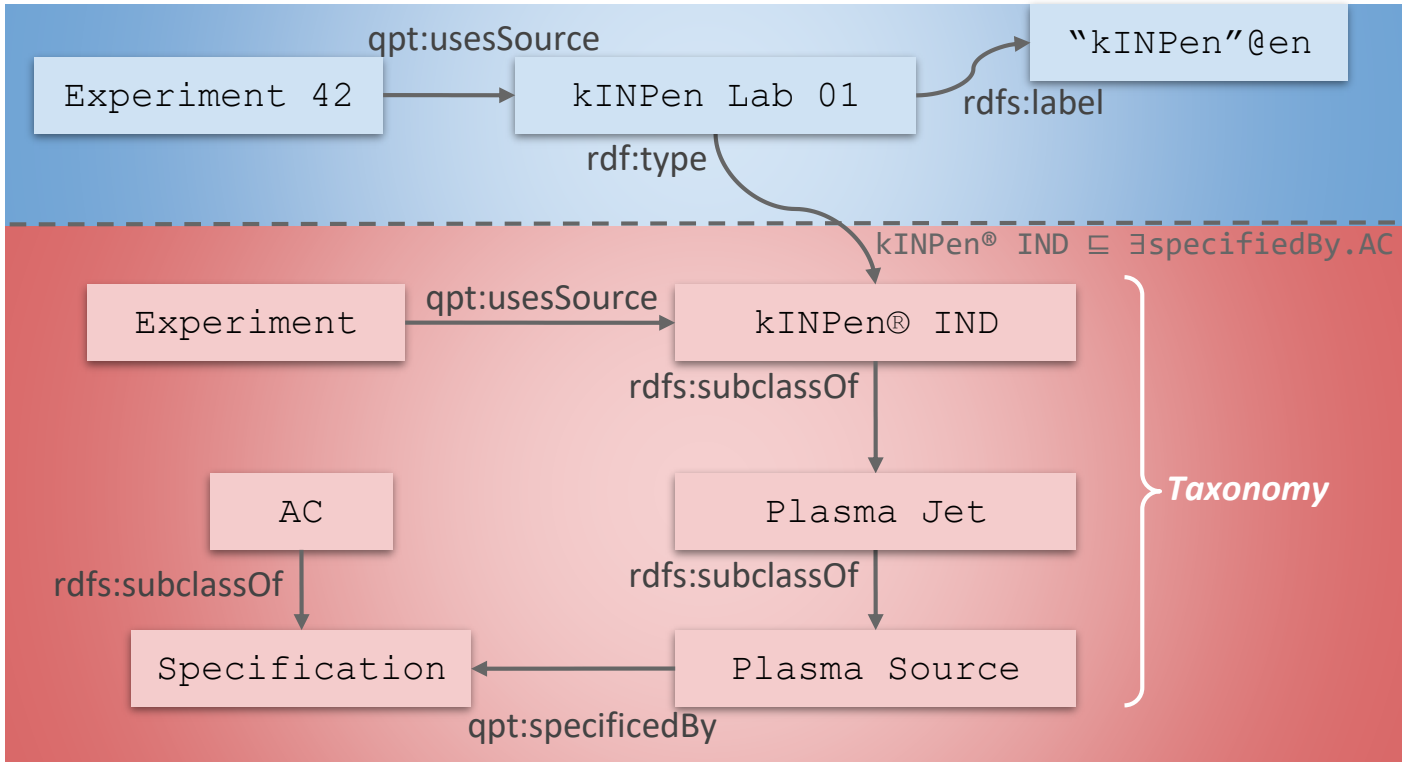


Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

What is an ontology?

An ontology is an explicit, **formal** specification of a shared conceptualization.

→ **formal knowledge representation**



Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Advantages

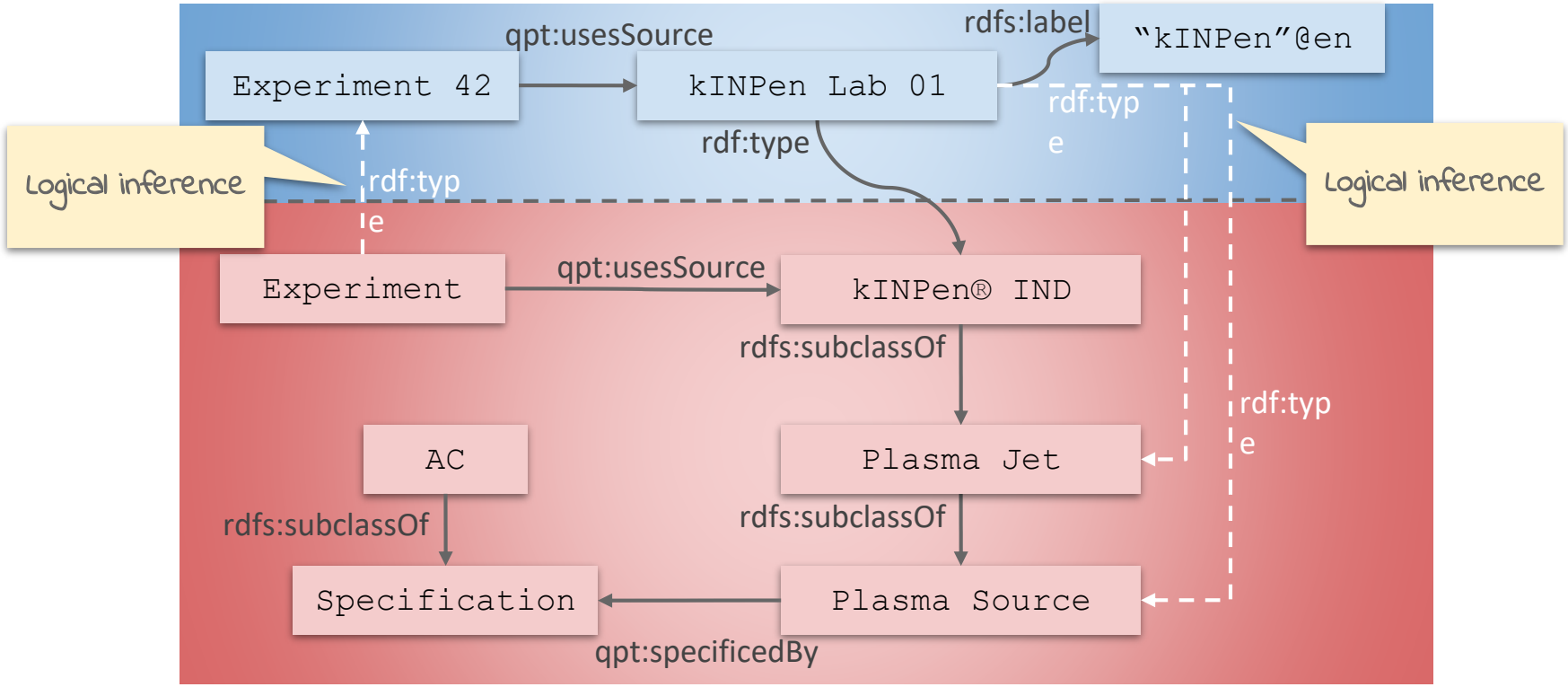
Ontologies

- Makes data FAIR
  - Findable
    - All classes, properties and instances have URIs
  - Accessible
    - Uses standards, like RDF, RDFS, OWL, SPARQL
  - Interoperable
    - uses a formal and broadly applicable language for knowledge representation
    - Ontologies are underspecified → easy to extend, connect to other ontologies
  - Reusable
    - well-documented and curated

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Advantages (II)

- Reasoning
  - Logical inference makes implicit knowledge explicit
  - Data can be checked for contradictions

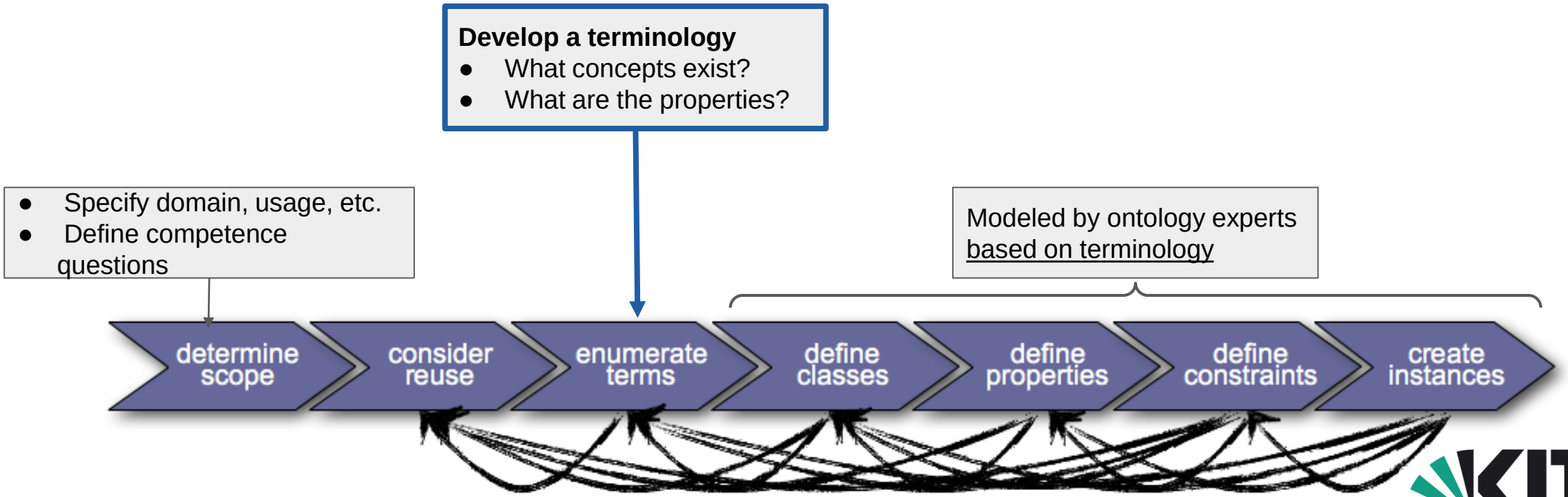


Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Development process

Ontology Development 101 *Noy, McGuinness, 2001 [4]*

- iterative process, constantly improving ontology



Research Data Management

Storing research data in a structured way with all required metadata.

- should be FAIR
- follow a common data schema

→ Knowledge Graphs, Ontologies

Community needs common schema

- Research is highly flexible
  → frequent changes
  → constant development effort necessary
- heterogenous interest of domain experts

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Community driven Development

Develop a terminology

- whole community has to agree on concepts, properties
- duplicates need to be identified
  → pure document based approaches are infeasible

Requirements

- Documentation of current development state
  - Already modeled concepts should be described in natural language
- Discussion platform
- Enable modelling of less complex elements by domain experts
- Easy linking towards concepts
- Small barrier to contribute
- Version Control System

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

MediaWiki

Why not use the most successful community driven project?

MediaWiki

- well-known among researchers
  - small barrier to contribute
- Discussion platform
- Articles to describe single concepts
- Version Control

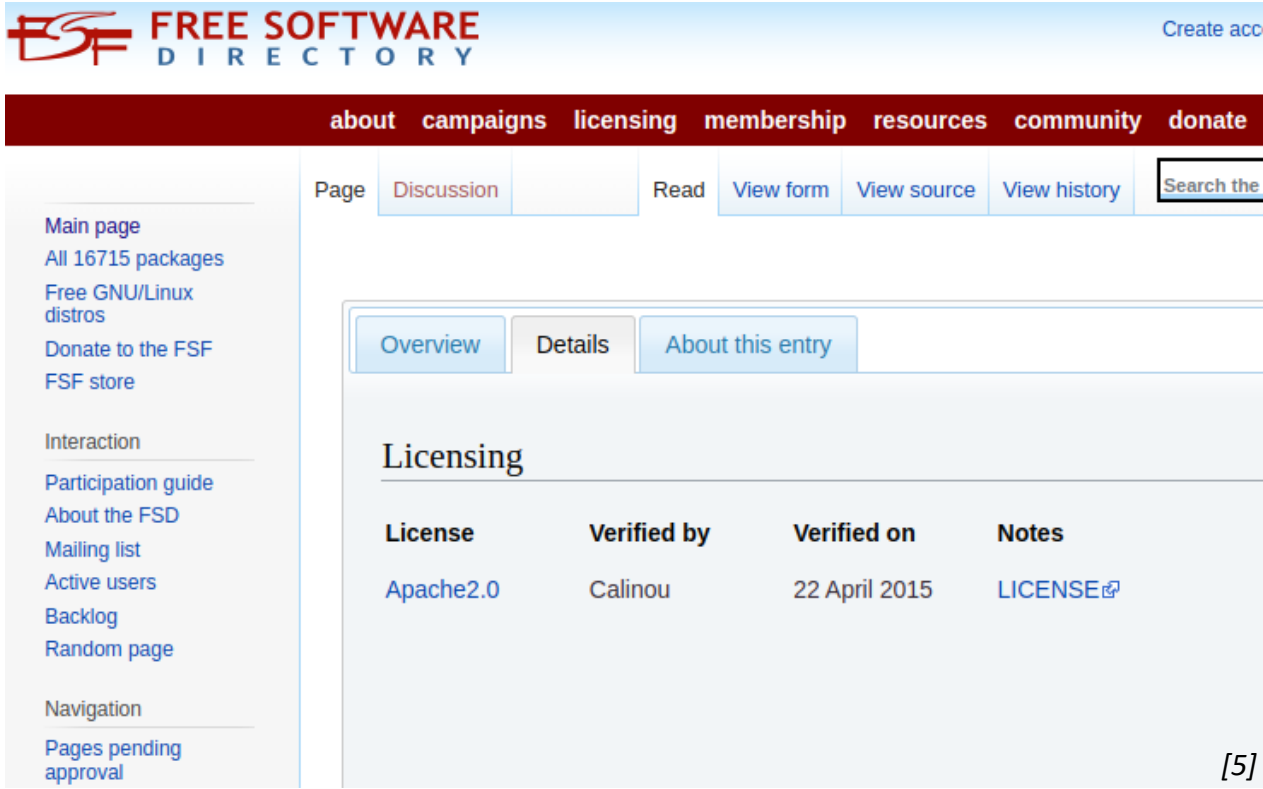
→ How to get the semantics?

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Semantic MediaWiki

MediaWiki Extension

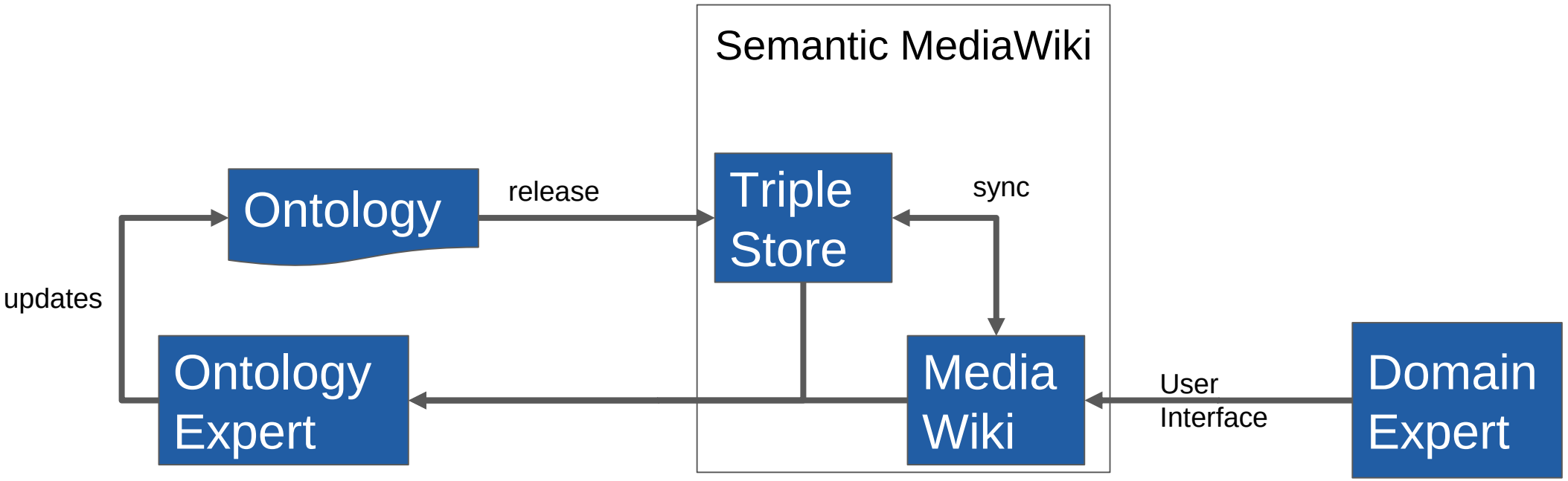
- enables storing and querying of RDF data in wiki pages
- simple wikitext based syntax
  - Defining RDF triple
    - `[[property::value]]`
    - `[[Category:PlasmaSource]]`
  - Querying
    - ```
      {{#ask:
          [[Category:PlasmaSource]]
          |?label }}
      ```
- Triple Store with RDF data



[5]

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.
      ```

Workflow



Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Interface

Category | Discussion

Category:PlasmaSource

Plasma Source

A plasma source is linked to Specification by the property specifiedBy.

Subcategories

This category has only the following subcategory.

P

- PlasmaJet

Category | Discussion

Editing Category:PlasmaSource

```
Plasma Source

A plasma source is linked to [[specifiedBy::Specification]]
by the property specifiedBy.
```

Conclusion

Key Takeaways

- Ontologies are a powerful tool for Research Data Management
- Semantic MediaWiki could be an easy way to involve the community

Discussion

- Are ontologies the right tool for your RDM?
  - Do you think the barrier to entry is to high?
- Would you contribute to an ontology?
  - Is using a MediaWiki an incentive to contribute?
  - What other incentives could be helpful?

Fabian Hoppe et al. SMW for community-wide RDM. FAIR Data in Plasma Science. 2022.

Thank you!

Contact
Fabian Hoppe
fabian.hoppe@fiz-karlsruhe.de
@fab_hop
https://www.fiz-karlsruhe.de/en/forschung/lebenslauf-und-publikationen-fabian-hoppe

References

[1] Hayden, M. (2014). The Price of Privacy: Re-Evaluating the NSA. The Johns Hopkins Foreign Affairs Symposium
[2] Ackoff, R. L. (1989). From data to wisdom. Journal of Applied Systems Analysis 15: 3-9
[3] INPTDAT, Decontamination
[4] Noy, N., & McGuinness, D. L. (2001). Ontology development 101. Knowledge Systems Laboratory, Stanford University, 2001.
[5] Free Software Directory, NeoVim