Metadata schema and ontologies for FAIR research data in plasma technology

M. M. Becker[1,*], St. Franke[1], F. Hoppe[2,3], D. Loffhagen[1], H. Sack[2,3], V. Skwarek[4], T. Tietz[2,3], S. Tschirner[4], L. Vilardell Scholten[1]

[1]Leibniz Institute for Plasma Science and Technology (INP), Felix-Hausdorff-Straße 2, 17489 Greifswald, Germany
[2]FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH, 76344 Eggenstein-Leopoldshafen, Germany
[3]Karlsruhe Institute of Technology (KIT), Institute AIFB, 76133 Karlsruhe, Germany
[4]Hochschule für Angewandte Wissenschaften Hamburg (HAW Hamburg), Berliner Tor 5, 20099 Hamburg, Germany

*E-mail: markus.becker@inp-greifswald.de

INTRODUCTION

- Steady increase of need for public storage of digital research data (policies of funders, publishers, and institutions, transparency of research)
- Essential factors for efficient re-use of data, e.g. for data driven science [1, 2]: findability (F), accessibility (A), interoperability (I), and re-usability (R) of research data → FAIR data principles
- Lack of common standards and tools to publish data according to these FAIR data principles in low-temperature plasma physics
- Modular metadata standards, ontologies and a plasma knowledge graph providing a basis for further community activities for unified research data management under development

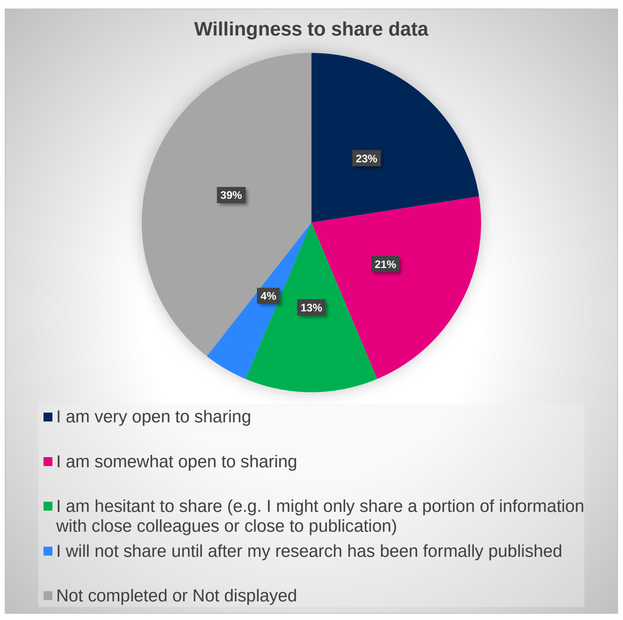
STATUS QUO

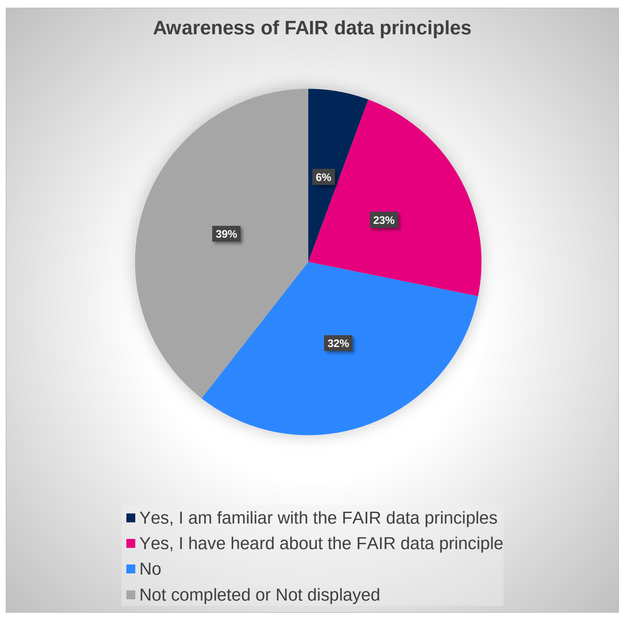
- International online survey to gather the s*tatus quo* of research data management practices, important factors and used standards in the low-temperature plasma physics community completed (answers: 71 total, 38 complete)
- Most responses representative for individual research groups
- Important conclusions from survey:
  - High willingness to share data
  - Need to increased awareness for possibilities and benefit of FAIR data
  - Requirement: transparent structures and standards for storage, documentation, processing, quality assurance, and publication of data

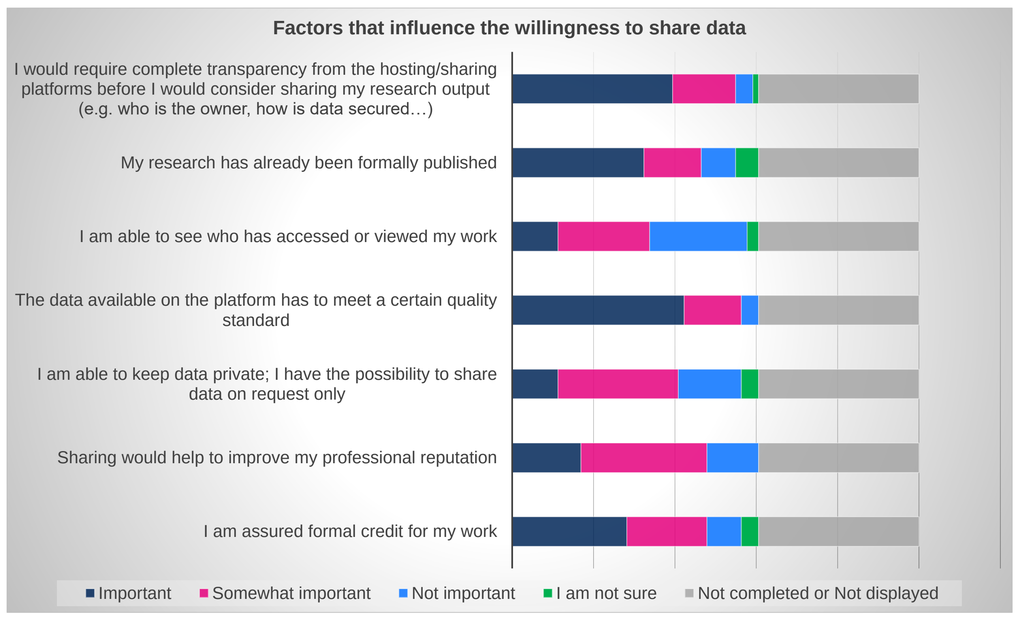
FAIR DATA PRINCIPLES

To be Findable	
F1	(meta)data are assigned a globally unique and persistent identifier
F2	data are described with rich metadata (defined by R1 below)
F3	metadata clearly and explicitly include the identifier of the data it describes
To be Accessible	
A1	(meta)data are retrievable by their identifier using a standardized communications protocol
A1.1	the protocol is open, free, and universally implementable
A1.2	the protocol allows for an authentication and authorization procedure, where necessary
A2	metadata are accessible, even when the data are no longer available
To be Interoperable	
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2	(meta)data use vocabularies that follow FAIR principles
I3	(meta)data include qualified references to other (meta)data
To be Reusable	
R1	meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1	(meta)data are released with a clear and accessible data usage license
R1.2	(meta)data are associated with detailed provenance
R1.3	(meta)data meet domain-relevant community standards

M. D. Wilkinson *et al., Scientific Data* **3** (2016) 160018.



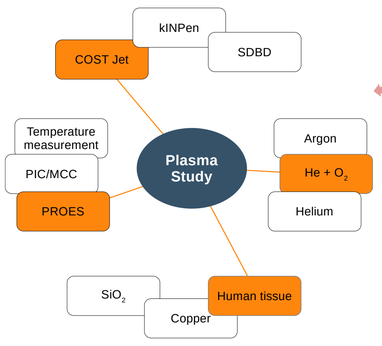


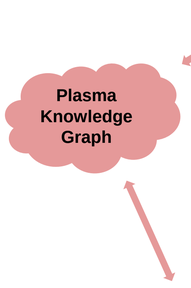


APPROACH

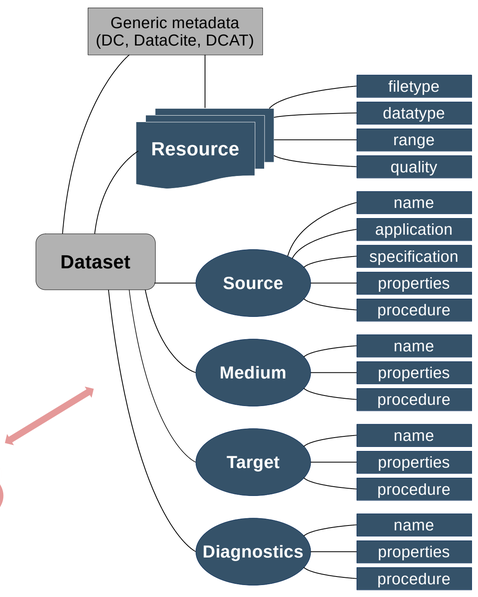
- Process oriented documentation of data on the basis of Plasma-MDS as general standard for all fields of plasma science and technology [3]
- Modularized extensions for specific methods and resources
- Common ontology and knowledge graph as basis for unified terminology, linking of (meta-) data and quality criteria

Components of plasma study



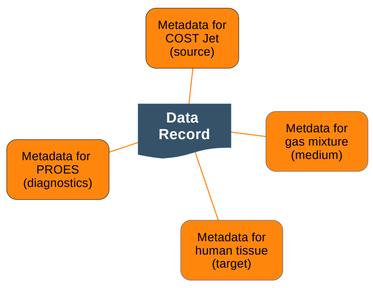


Plasma Metadata Schema



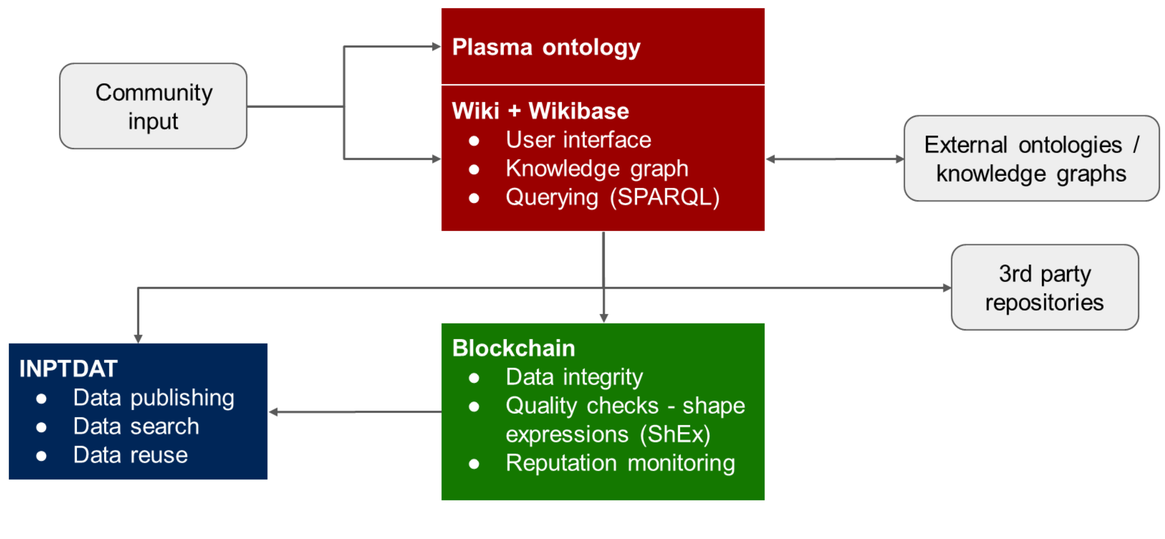
St. Franke *et al.*, arXiv:1907.07744, 2020.

Metadata modules for specific study



PLANNED INFRASTRUCTURE

- Extension of INPTDAT data platform (https://www.inptdat.de) by Wikibase system for community driven plasma knowledge graph, and blockchain infrastructure
- Open interfaces for integration of the plasma knowledge graph into external services and tools, e.g. data bases, research data repositories, and electronic lab books



BENEFIT

- Visibility, transparency and reproducibility of scientific results
- Validation of quantitative results
- Less reproduction of same data
- Support of standardization
- New findings by meta analyses and data driven research
- Building confidence in data

References:

[1] M. D. Wilkinson *et al., Scientific Data* **3** (2016) 160018.
[2] GO-FAIR: FAIR Principles, https://www.go-fair.org/fair-principles.
[3] St. Franke *et al.*, arXiv:1907.07744, 2020.

SPONSORED BY THE

Federal Ministry of Education and Research

The German Federal Ministry of Education and Research (BMBF) funded this work under grant marks 16QK03A, 16QK03B, and 16QK03C.