Development of a research data management infrastructure for plasma technology

Markus Becker

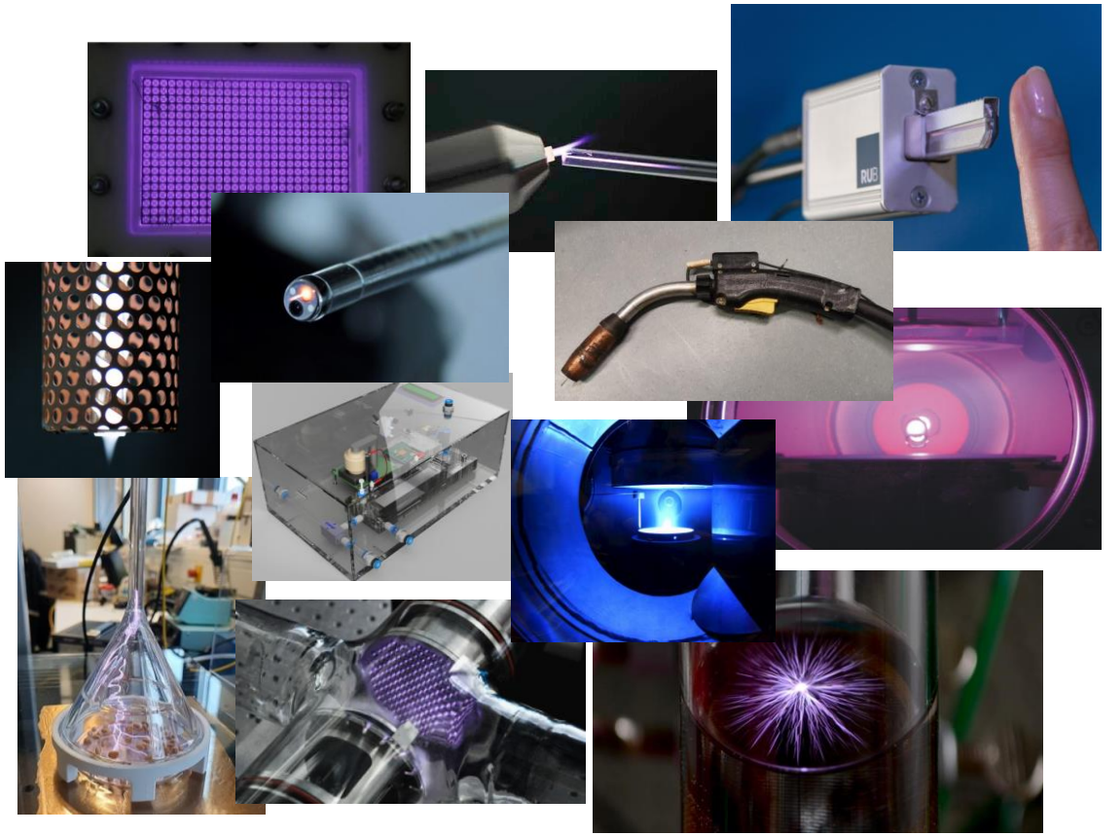
Round-Table „Forschungsdatenmanagement“
2021-05-20, virtual

Leibniz Gemeinschaft

FROM IDEA TO PROTOTYPE

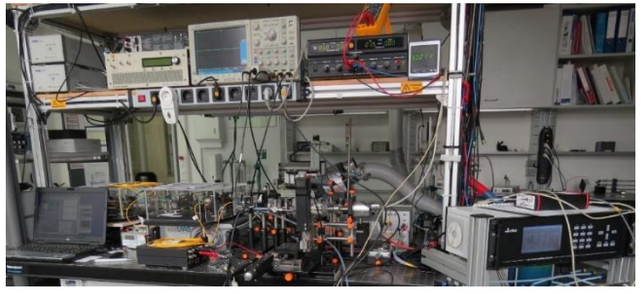
Challenges for research data management

Diversity of plasma sources



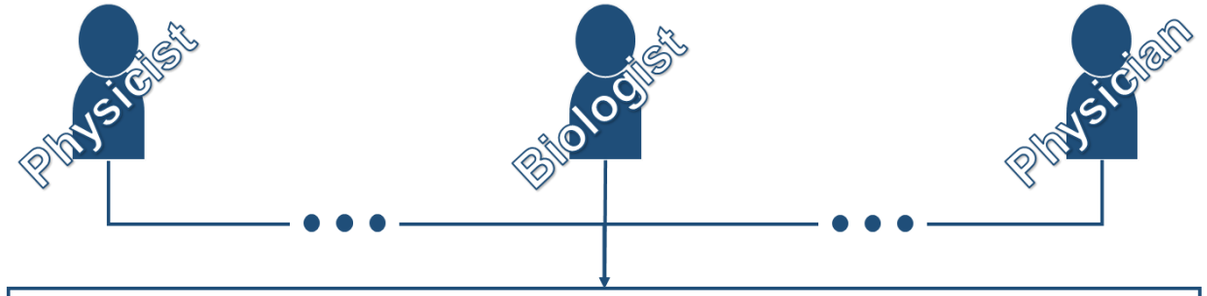
Diversity of methods

- Method
  - **Experimental method**
    - Set of methods
      - DinSpec
    - Electrical Method
    - Surface diagnostics
      - Ellipsometry:
    - Imaging
      - High-speed photography
      - High-speed imaging (fast imaging)
      - Thermal imaging
      - Infrared thermography (IRT)
      - Scanning electron microscopy (SEM)
      - Atomic Force Microscope (AFM)
      - Scanning Tunneling Microscope (STM)
    - Mechanical Method
      - Profilometry
      - Probe measurements
        - Langmuir probe measurements
        - Multipole resonance probe measurement (MRP)
        - Atom Probe Tomography (APT)
    - Spectroscopy
      - Mass spectrometry (MS)
      - Coherent anti-Stokes Raman spectroscopy (CARS)
      - Laser absorption spectroscopy (LAS)
      - Optical Emission Spectroscopy (OES)
      - X-ray photoelectron spectroscopy (XPS)
      - Fourier-Transform Infrared Spectroscopy (FTIR)
      - Raman spectroscopy
      - Energy-dispersive X-ray spectroscopy (EDX)
  - **Theoretical / simulation methods**
    - Hybrid method
    - First principle methods
    - Kinetic methods
      - Particle-in-cell simulation (PIC)
    - Hydrodynamic methods (fluid methods)
      - Fluid-Poisson model
    - Reaction kinetics (chemical kinetics)
      - Collisional-radiative model (CRM)

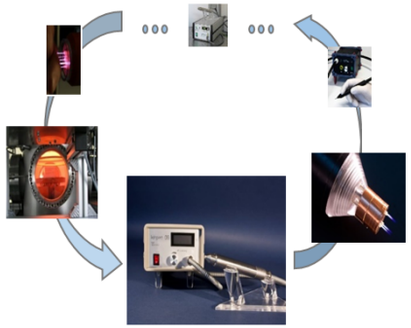


Interdisciplinarity of research fields like plasma medicine

InPT-Dat concept (project at INP from 2017 to 2019)



InPT-Dat data platform



- Storage and sharing of research data
- Reference to corresponding publications
- Content indexing and full-text search
- Direct linking and visualisation of data records
- Directory of plasma sources and applications

Benefit

- Version-safe long-term archiving of research data according to the guidelines of good scientific practice.
- Simplified reuse of interdisciplinary research data, especially for researchers from other fields.
- Merging of heterogeneous research data from different fields of science → generation of new scientific findings.

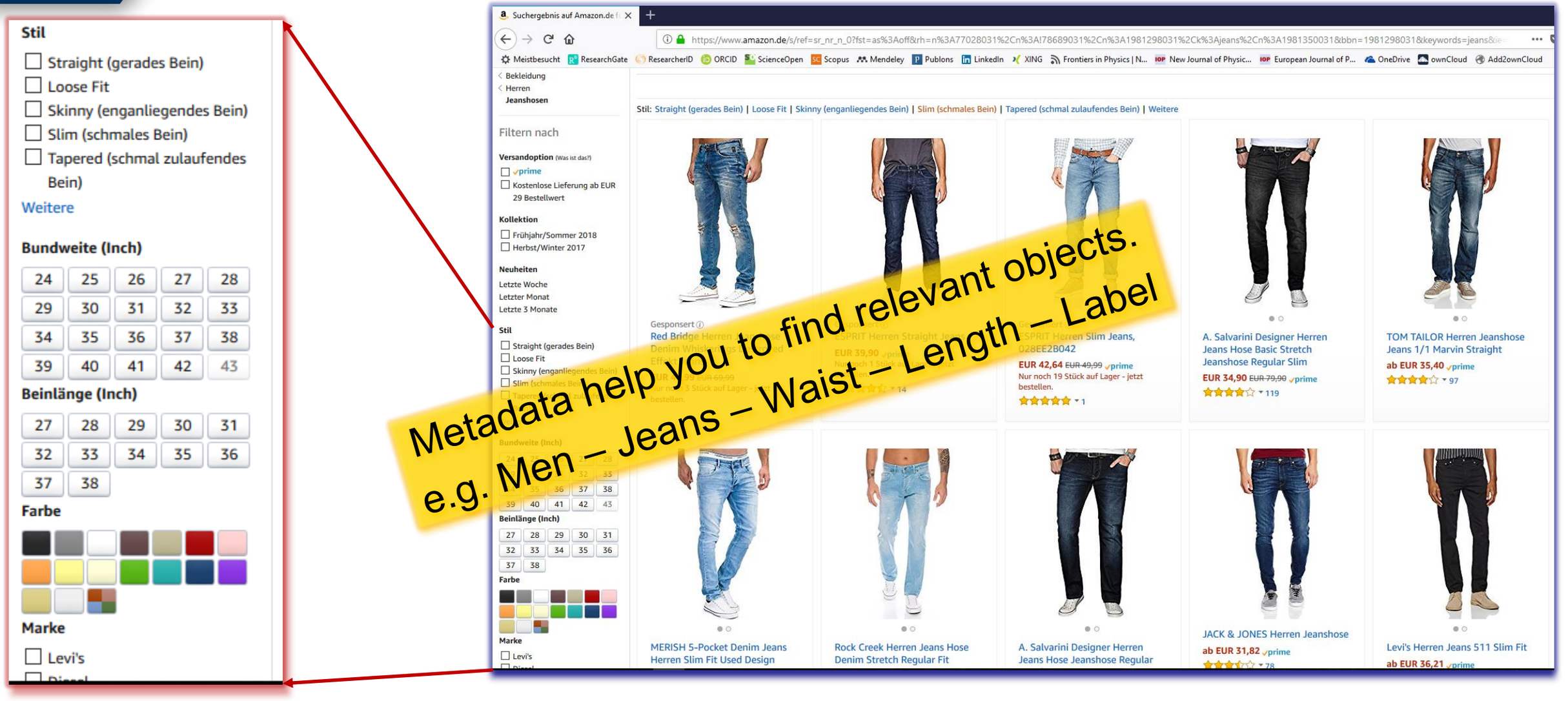
SPONSORED BY THE

Federal Ministry of Education and Research

Approach:

1. Subject specific metadata schema
2. Institutional data platform

What are Metadata? What are they for?



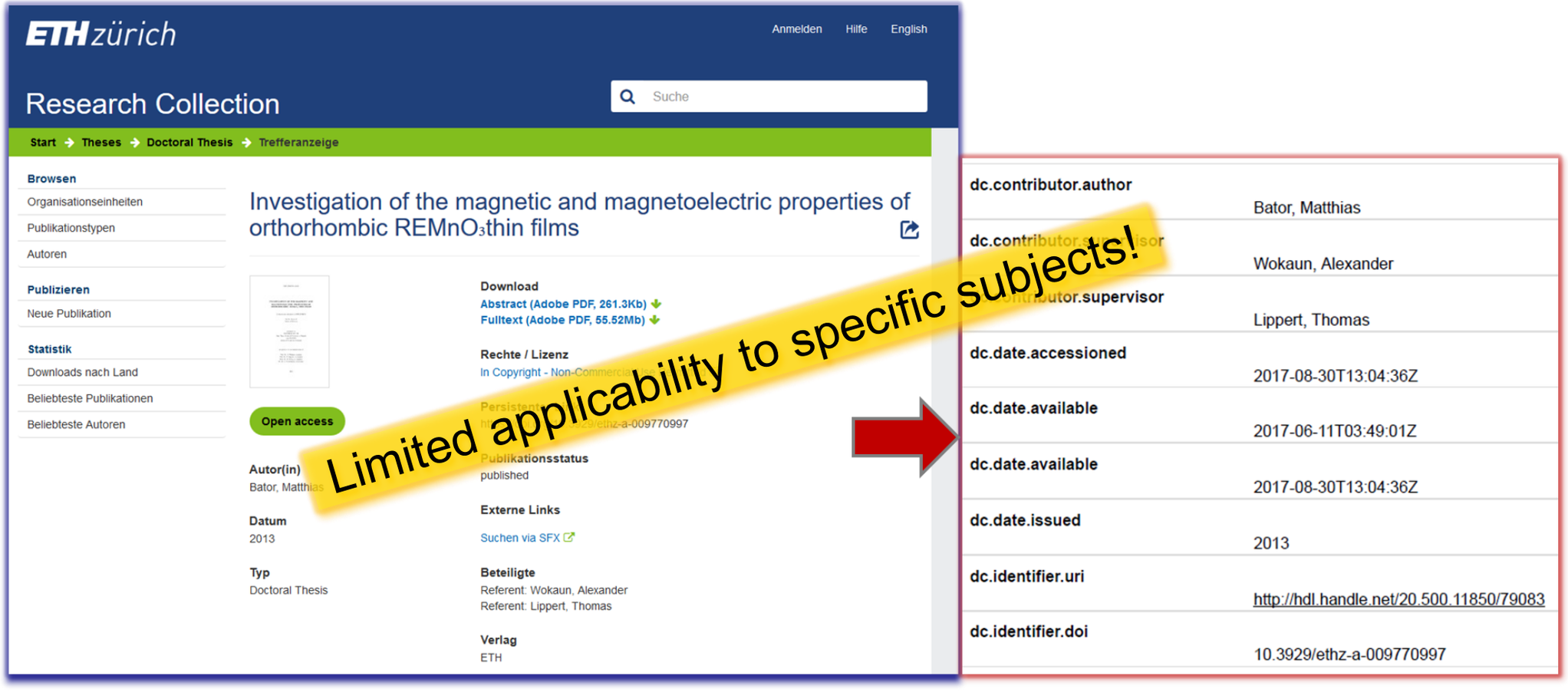
What are Metadata? What are they for?

- Data about data / object.
- Compressed information.
- Help to filter objects out of a pool.

- Metadata answer core questions about an object, like
  - Who? – Author/Creator/Institution
  - What? / Why? – Title/Abstract
  - When? – Date submitted/accepted/published
  - Where? – Journal/Identifier (DOI/URL)

Dublin Core: The core standard of metadata

1. Identifier
2. Format
3. Type
4. Language
5. Title
6. Subject
7. Coverage
8. Description
9. Creator
10. Publisher
11. Contributor
12. Rights
13. Source
14. Relation
15. Date



Subject specific metadata standards

- http://www.dcc.ac.uk/resources/subject-areas/physical-science

Physical Science

Materials Science Geography Geology Geoscience Crystallography Solar physics Astrophysics Molecular biology Multi-disciplinary Biochemistry Meteorology Space science Remote Sensing Chemistry Nuclear and Particle Physics Physics Bioinformatics Astronomy

Metadata Standards

AVM - Astronomy Visualization Metadata
A standard defining discovery metadata for fully rendered astronomical imagery.

CIF - Crystallographic Information Framework
An extensible standard file format and set of protocols for the exchange of crystallographic and structured data.

CSMD-CCLRC Core Scientific Metadata Model
A study-data oriented model that captures high-level information about scientific studies and the they produce, primarily tailored for the physical sciences.

FITS - Flexible Image Transport System
Used by the astronomy community to originally describe telescope images, but is now a family to describe multi-dimensional data including spatial, spectral and temporal dimensions and the distortions that may be present.

No standard exists for plasma physics

- IVOA - Resource Metadata for the Virtual Observatory

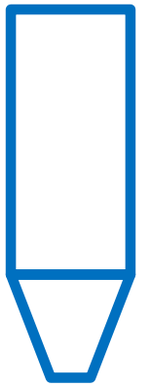
Collection and service content metadata

Facility	Apache Point Observatory, Sloan 2.5-m Telescope
Instrument	Five-band clocked CCD camera
Coverage.Spatial	PositionInterval FK5 145.17 –1.25 235.9 1.25 PositionInterval FK5 250.71 52.15 267.0 66.29 PositionInterval FK5 350.43 –1.25 359.99 1.17 PositionInterval 0.0 –1.25 56.37 1.17
Coverage.RegionOfRegard	0.0001
Coverage.Spectral	Optical
Coverage.Spectral.Bandpass	u', g', r', i', z'
Coverage.Spectral.MinimumWavelength	400.e-9
Coverage.Spectral.MaximumWavelength	850.e-9
Coverage.Temporal.StartTime	1999-12-25
Coverage.Temporal.StopTime	2001-07-15
Coverage.Depth	3.e-6
Coverage.ObjectDensity	6.e4
Coverage.ObjectCount	2.e7
Coverage.SkyFraction	0.01

Development of a general concept

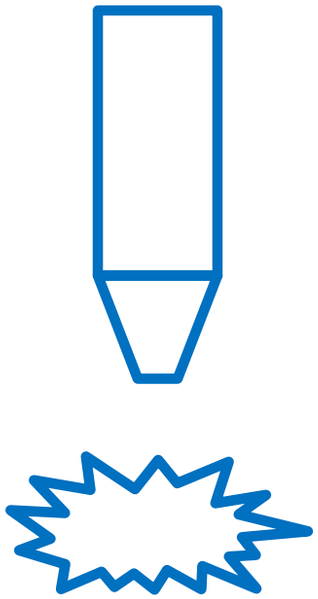
- Plasma
  - … is generated by a plasma **source**
  - … is operated with/in a **medium**
  - … is sometimes used to treat a **target**
- **Diagnostics** (experimental and/or computational methods) are used to study the plasma, medium and/or target
- Stored research data (**resources**) are often useless without information about the whole process
- Application and specification of the plasma would help to find relevant data sets

Plasma source (APPJ, COST-Jet, kINPen)



Metadata field
plasma.source.name
plasma.source.application
plasma.source.specification
plasma.source.properties
plasma.source.procedure

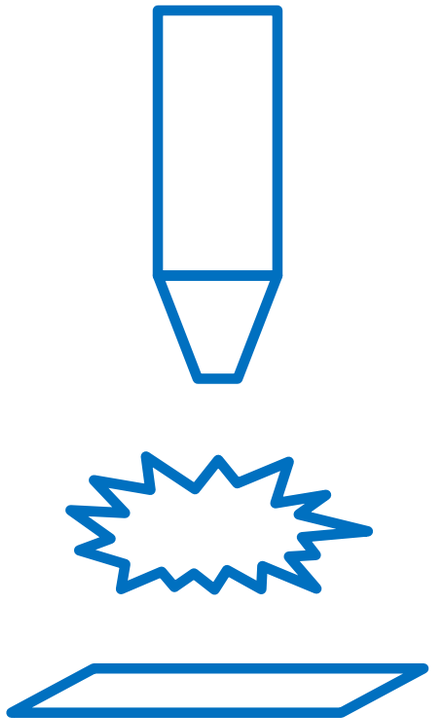
Plasma medium (air, Ar, O2)



Metadata field
plasma.source.name
plasma.source.application
plasma.source.specification

Metadata field
plasma.medium.name
plasma.medium.properties
plasma.medium.procedure

Plasma target (silicon wafer, E. coli, distilled water)

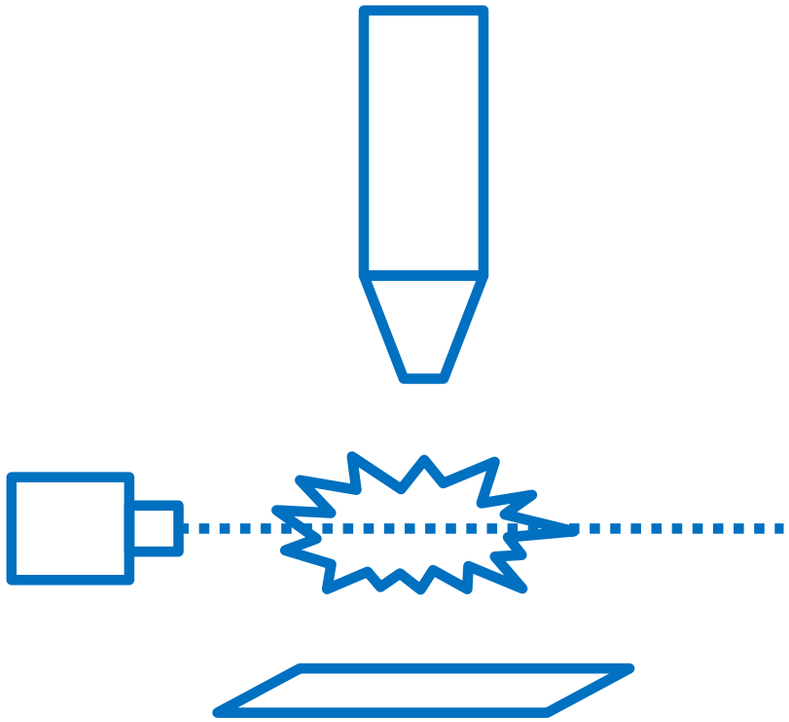


Metadata field
plasma.source.name
plasma.source.application
plasma.source.specification

Metadata field
plasma.source.name

Metadata field
plasma.target.name
plasma.target.properties
plasma.target.procedure

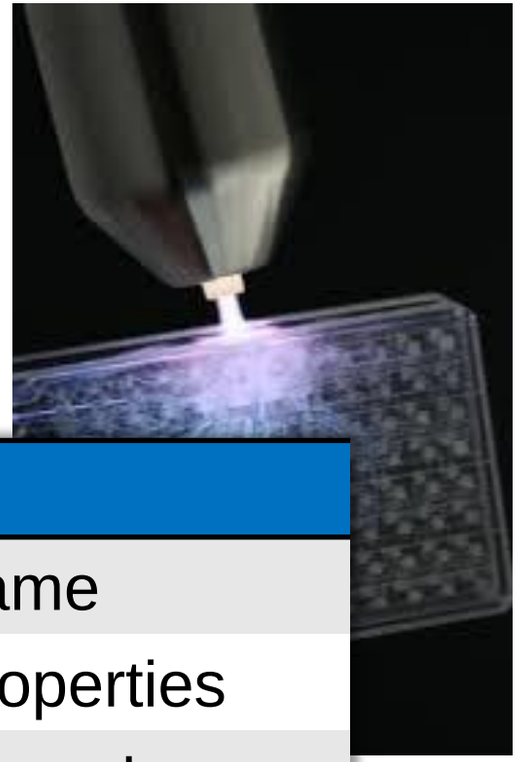
Diagnostics / modelling / simulations (OES, XPS, PIC-MCC)



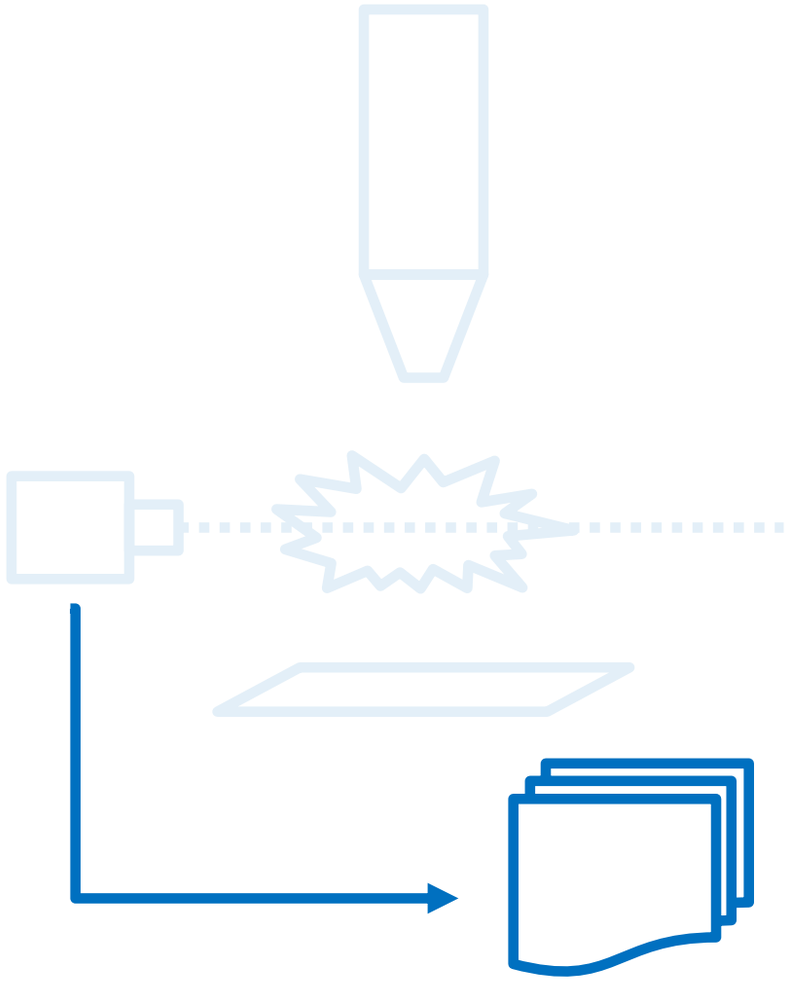
Metadata field
plasma.source.name
plasma.source.application
plasma.source.specification

Metadata field
plasma.target.procedure

Metadata field
plasma.diagnostics.name
plasma.diagnostics.properties
plasma.diagnostics.procedure

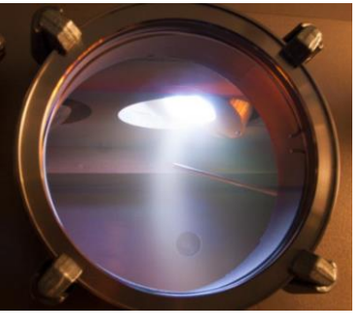


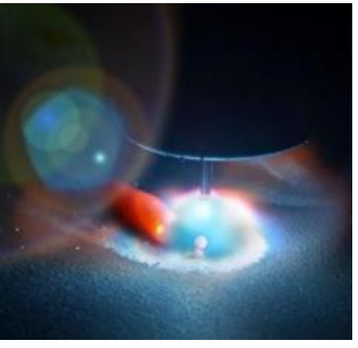
Resource (data file)



Metadata field
plasma.source.name
plasma.source.application
plasma.source.specification
Metadata field
Metadata field
Metadata field
plasma.resource.filetype
plasma.resource.datatype
plasma.resource.range
plasma.resource.quality

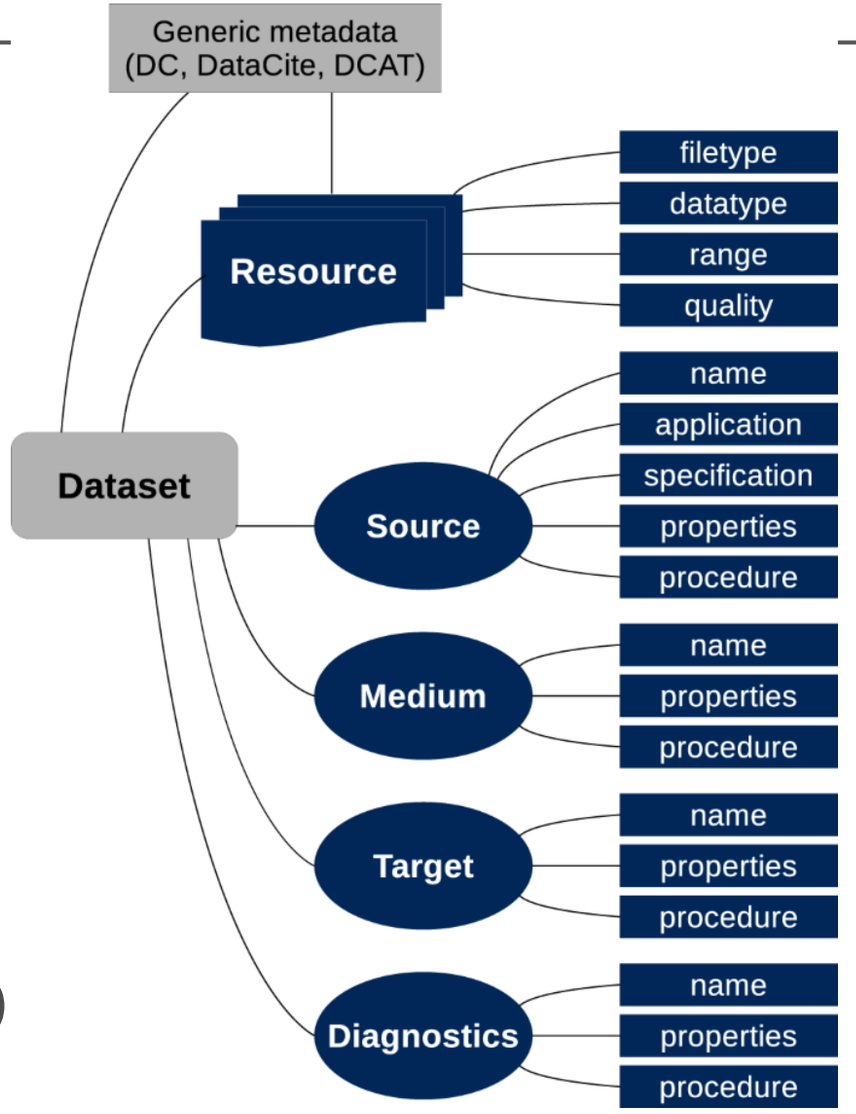
Plasma metadata schema





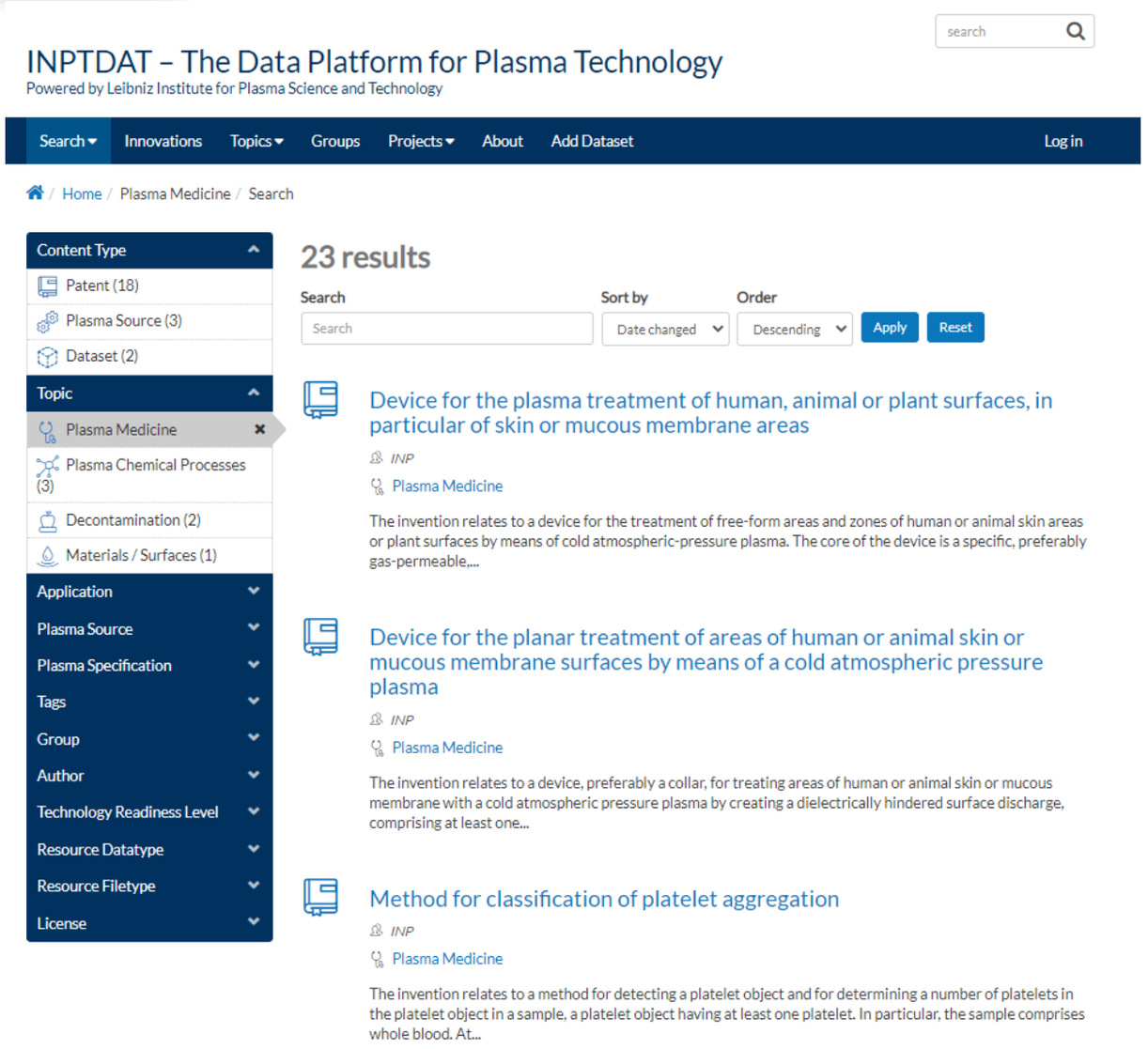
Plasma-MDS

- Disciplinary metadata schema for plasma science
- Metadata fields for description of
  - plasma source
  - plasma medium
  - plasma target
  - diagnostics / modelling / simulations
  - resources (data)
- Research data repositories implementing Plasma-MDS at INP (https://www.inptdat.de) and RUB (https://rdpcidat.rub.de)



Franke, St. *et al.*, *Sci. Data* **7**, 439 (2020)

INPTDAT data platform



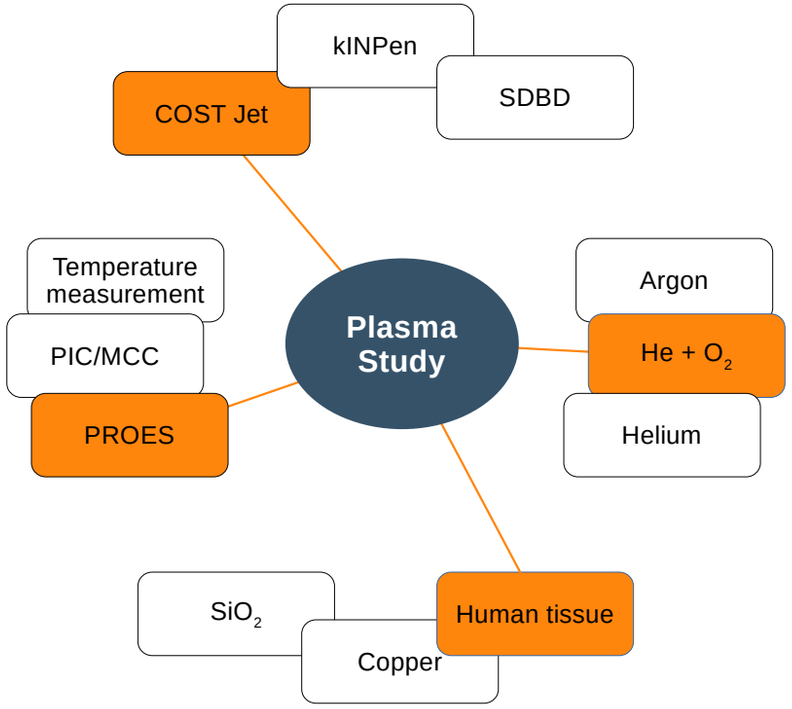
- Data publications with DOI for linking of journal paper and data set

> 17. Zhang G, Goett G, Uhrlandt D, Lozano Ph, Sharma R (2020). A simplified voltage model in GMAW, INPTDAT https://doi.org/10.34711/inptdat.146

- Linking of data sets, patents and plasma sources
- Plasma-MDS for uniform annotation of data
- Facetted search for data filtering
- External data sets can be included in the catalog (automated exchange of metadata between related repositories in preparation)
- The source code is on Github: https://github.com/plasma-mds/inptdat-platform

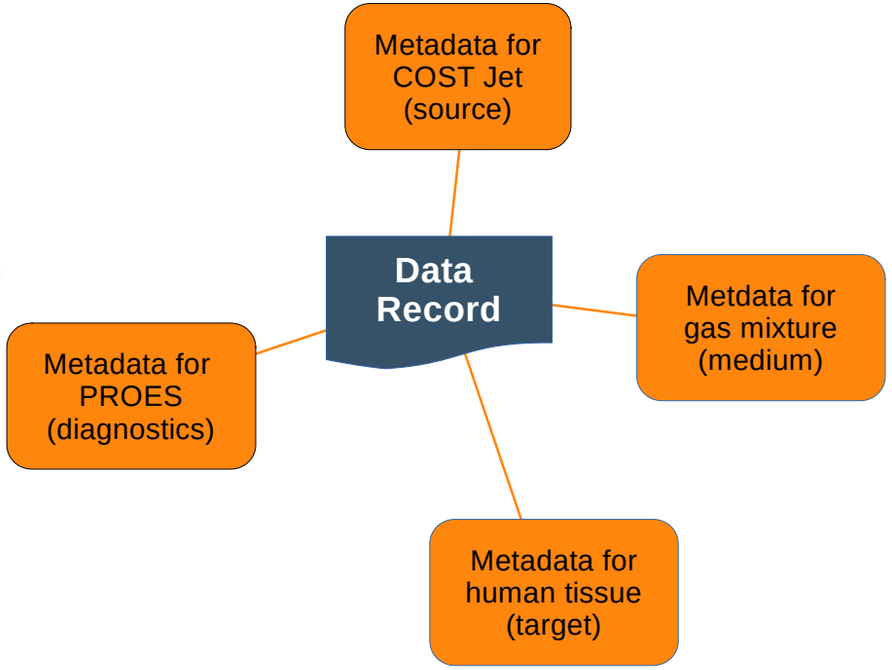
Next step: modular refinement of metadata

Metadata modules for different devices, methods, substrates etc.



Selection according to given requirements

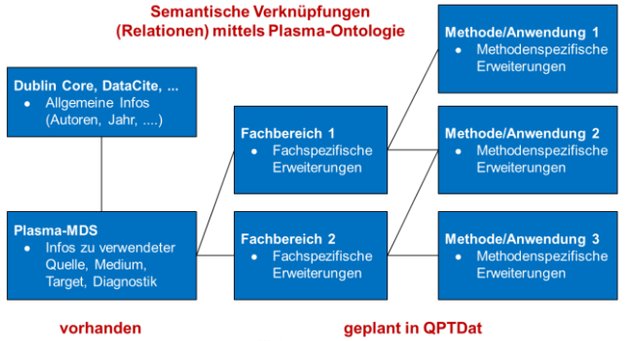
Individual compilation of data documentation templates



QPTDat concept (joint project from 2019 to 2022)

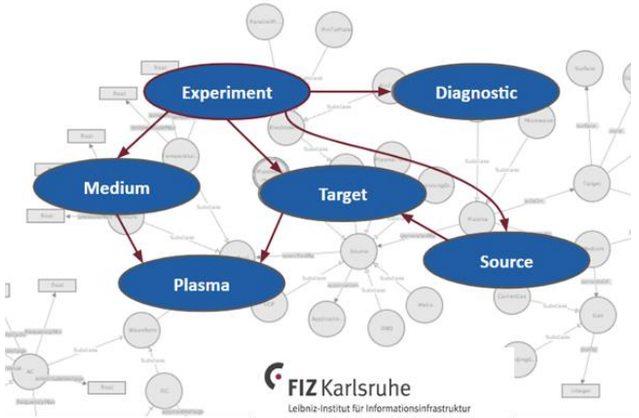
INP FROM IDEA TO PROTOTYPE

Metadata modules and quality criteria for reference cases



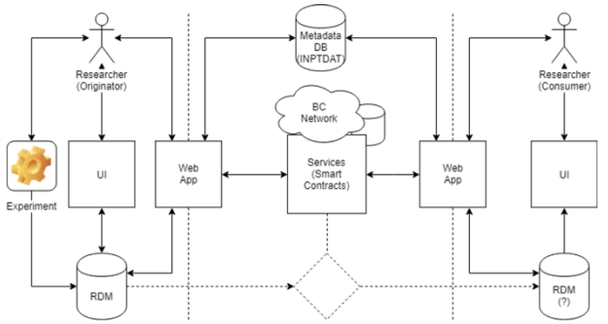
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Ontology and knowledge graph for semantic linking



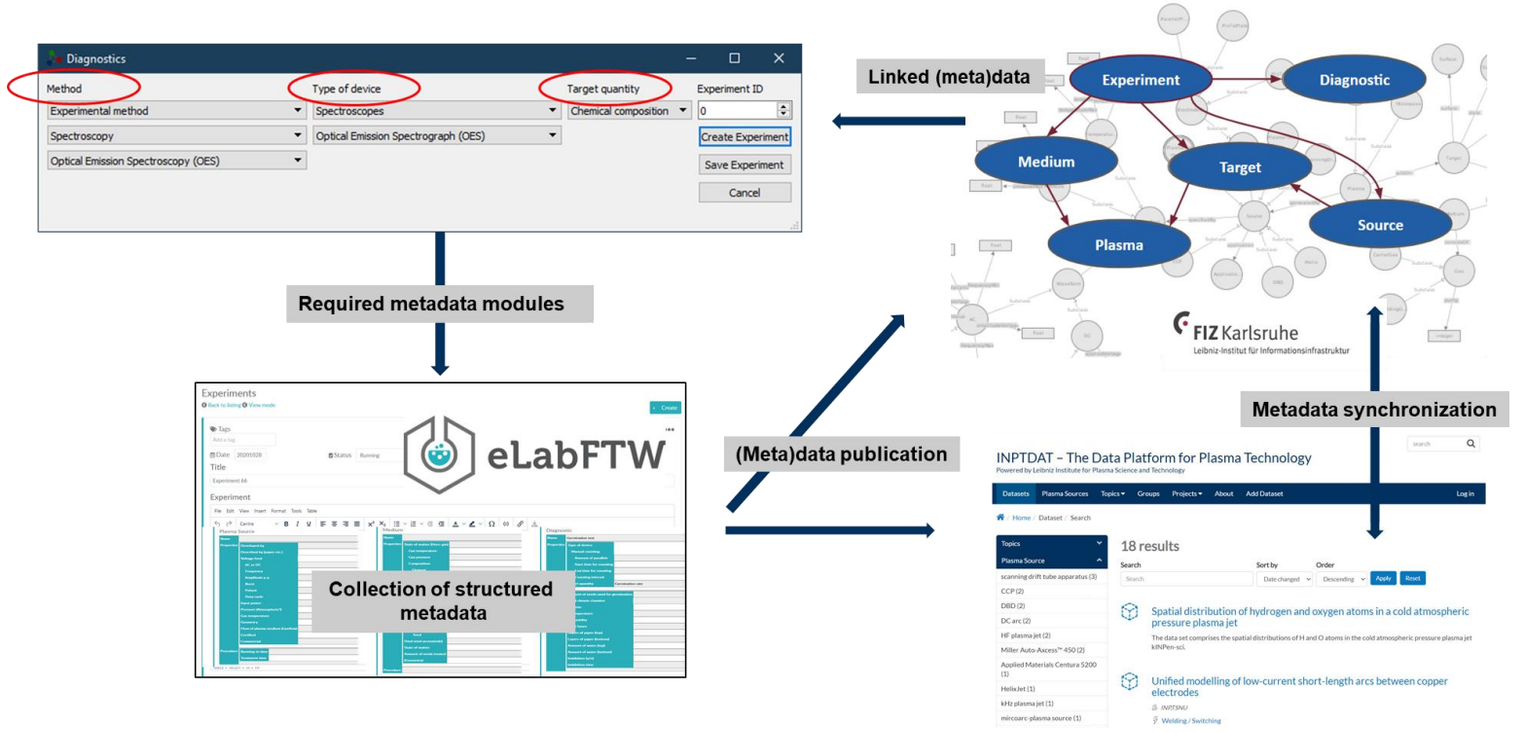
HAW HAMBURG

Blockchain for data certification and reputation monitoring



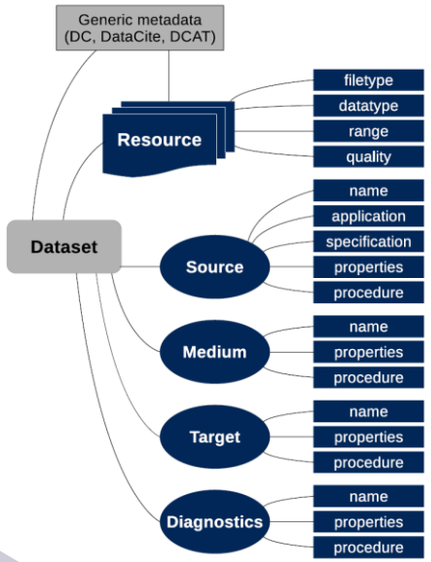
Community involvement

- Github repository as a starting point for broader development of joint concepts: https://github.com/plasma-mds/plasma-metadata-schema
- Development of metadata modules and related templates for electronic lab notebooks (eLabFTW)
- The knowledge graph will provide the basis for automated processes and simplified data annotations in the long-term.



Conclusion and outlook

Development of metadata schema and data platform



https://www.inptdat.de

https://rdpcidat.rub.de

InPT-Dat

QPTDat

NFDI4Phys

Knowledge graph + Blockchain for implementation of quality criteria and RDM workflows

Community wide RDM tools and standards

Contact

Leibniz Institute for Plasma Science and Technology
Address: Felix-Hausdorff-Str. 2, 17489 Greifswald
Phone: +49 - 3834 - 554 3821, Fax: +49 - 3834 - 554 301
E-mail: markus.becker@inp-greifswald.de, Web: www. leibniz-inp.de