

FAIR Data Commons / Essential Services and Tools for Metadata Management Supporting Science

Rainer Stotzka, Gulzaure Abdildina, Rossella Aversa, Nicolas Blumenröhr, Sabine Chelbi, Leonhard Duda, Felix Ernst, Laura Frank, Germaine Götzelmann, Volker Hartmann, Maximilian Inckmann, Thomas Jejkal, Vandana Jha, Reetu Joseph, Philipp Ost, Andreas Pfeil, Yusra Shakeel, Mehmet Soysal, Danah Tonne, Philipp Tögel, Elias Vitali

Karlsruhe Institute of Technology
www.kit.edu
rainer.stotzka@kit.edu



Service Components & Tools

KIT Data Manager

It is a Research Data Repository Platform consequently following state-of-the-art recommendations and standards for FAIRly managing research data.

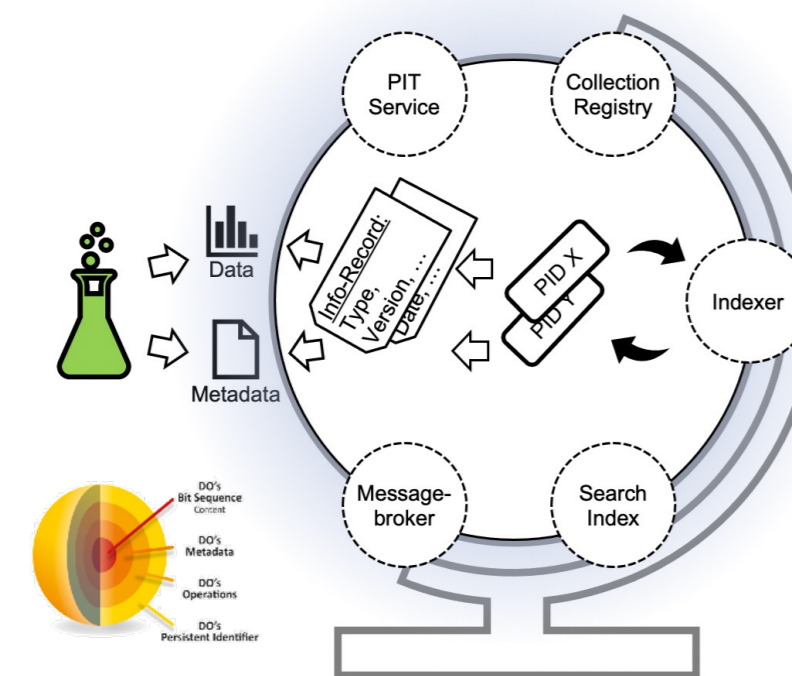
- Platform to build-up community repositories
- Source code: <https://github.com/kit-data-manager>
- Documentation: <https://kit-data-manager.github.io>



FAIR Digital Object Lab

It allows creation, modification, and validation of FAIR DOs. This out-of-the-box infrastructure based on Docker allows to gain first experiences with FAIR DOs, either using real PIDs or with sandboxed PIDs for testing purposes.

- Source code: <https://github.com/kit-data-manager/FAIR-DO-Lab>



Typed PID Maker

It provides a service for dealing with PID Information Types and Kernel Information Profiles according to the recommendations of the Research Data Alliance. The Typed PID Maker is one of the main building blocks for creating FAIR DOs and is a central part of the FAIR DO Testbed.

- Source code: <https://github.com/kit-data-manager/pit-service>

Web Annotation Protocol Server

An annotation is extra information associated with a particular part of research data. In contrast to traditional metadata, it may contain contradictory and dynamic information from various sources. The Web Annotation Protocol Server is an implementation of the W3C Web Annotation Data Protocol standard for preserving and managing annotations of research data.

- Source code: <https://github.com/kit-data-manager/wap-server>

ro-crate-java

Java library to create, validate, and to modify research packages (RO-Crates). These packages allow to describe research data machine-readable as well as human-readable, making it FAIR.

- Source code: <https://github.com/kit-data-manager/ro-crate-java>

- Source code: <https://github.com/kit-data-manager/ro-crate-benchmarks>

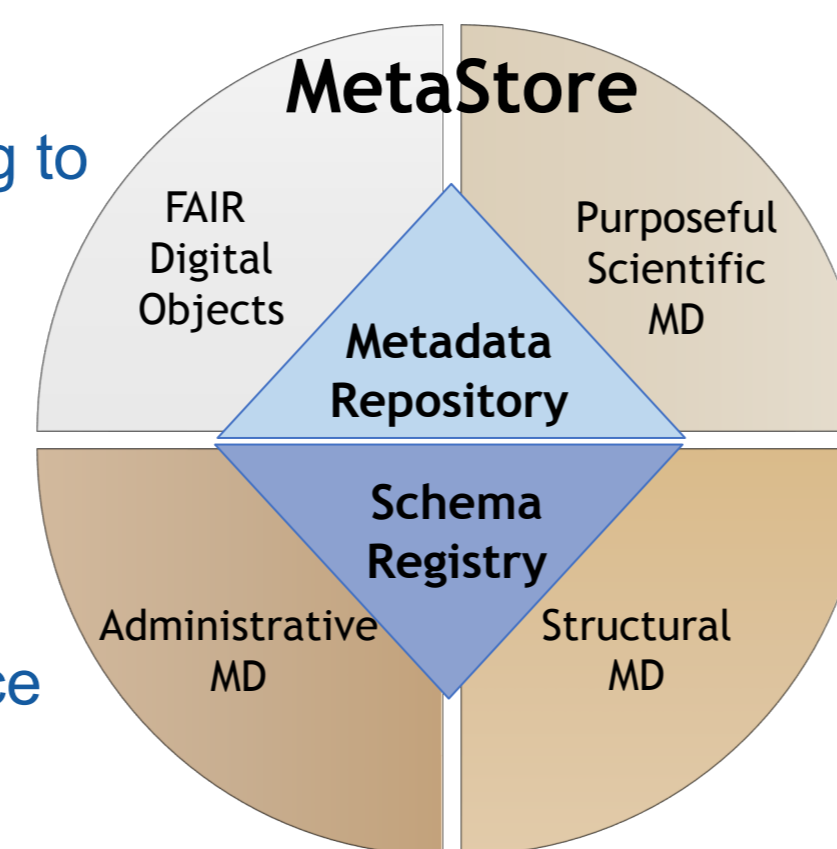


MetaStore

It is a metadata repository system allowing to manage huge collections of metadata documents and schemas.

- Millions of metadata documents
- Supports arbitrary XML and JSON schemas
- Metadata Editor as basic user interface
- OAI-PMH support

- Source code: <https://github.com/kit-data-manager/metastore2>



Metadata Repository Turntable

It provides a common API based on DOIP for various metadata repository systems.

Collaboration with

Source code: <https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3>

Documentation: <https://nfdi4ing.pages.rwth-aachen.de/s-3/s-3-3/turntable-interface>

Interface implementation: <https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/metadatabus>

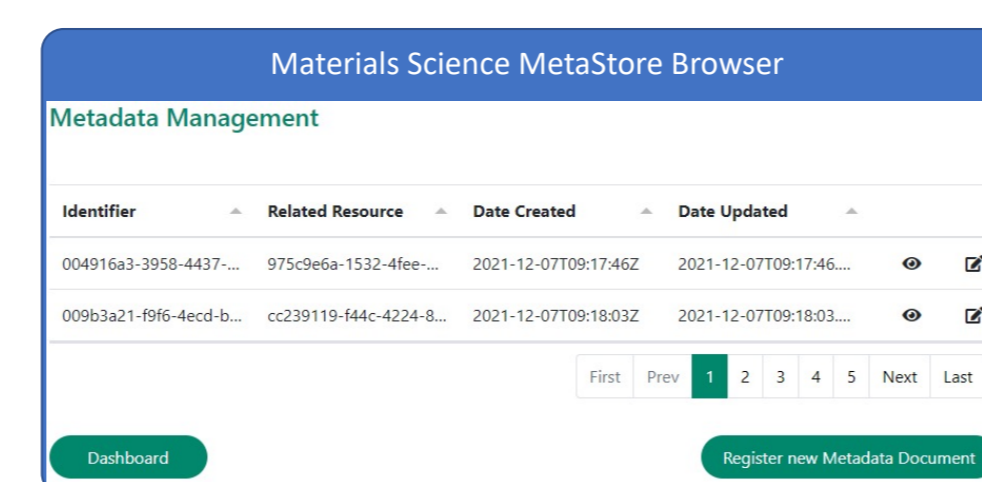
Testing client: <https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/client4metadatabus>

Metadata Editor

JavaScript library allowing to generate web forms and validate metadata in an intuitive and generic way. It builds a bridge between technical services and researchers.

Applications in materials science, nano science, digital humanities.

- Source code: <https://github.com/kit-data-manager/metadatabus>



Integration of the Metadata Editor as basic user interface (example of MetaStore). It easily allows developers to create graphical frontends for existing, machine-actionable services managing (structured) metadata.

Collection Registry

Implementation of the Collection API proposed by the RDA recommendation on Research Data Collections to create and manage collections of arbitrary digital resources. It allows to compile virtual collections of heterogeneous research data from various data resources.

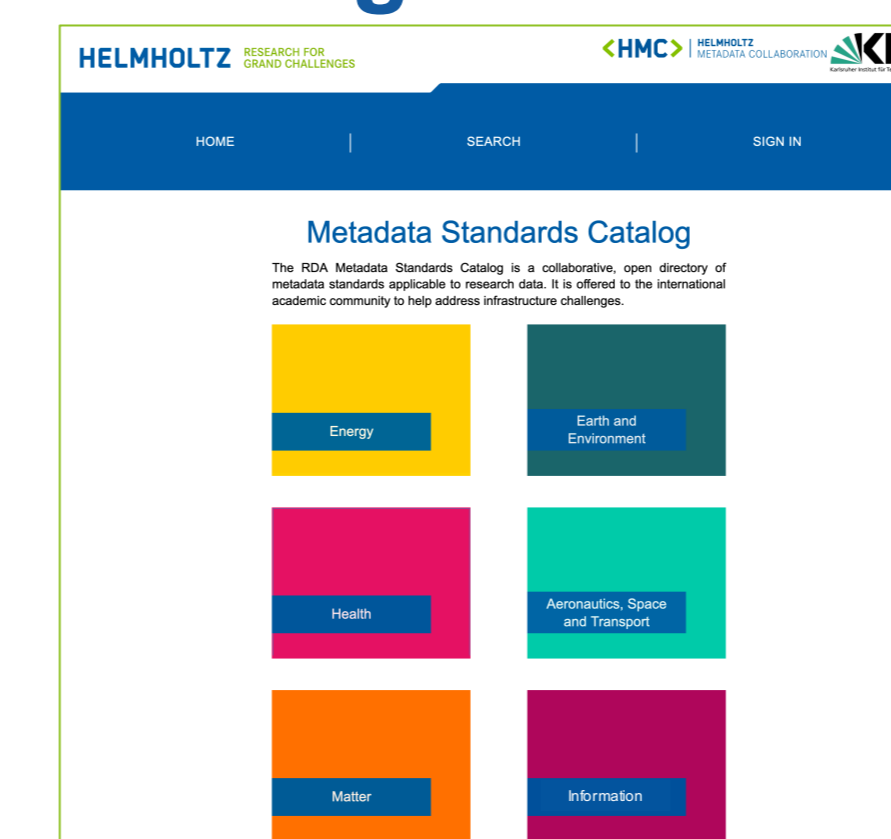
- Source code: <https://github.com/kit-data-manager/collection-api>

Vocabulary Service EVOKS

EVOKS allows to create, edit, curate, and manage SKOS vocabularies collaboratively in an easy manner.

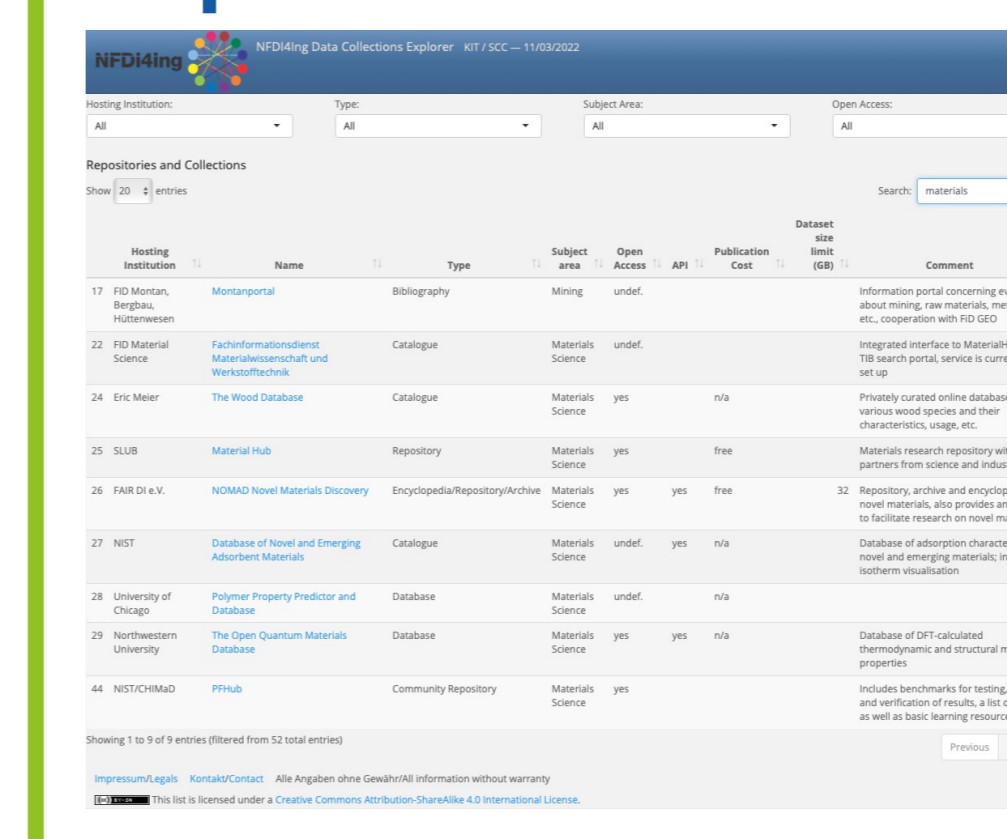
Information Systems

Metadata Standards Catalog



- Source code: <https://msc.datamanager.kit.edu>

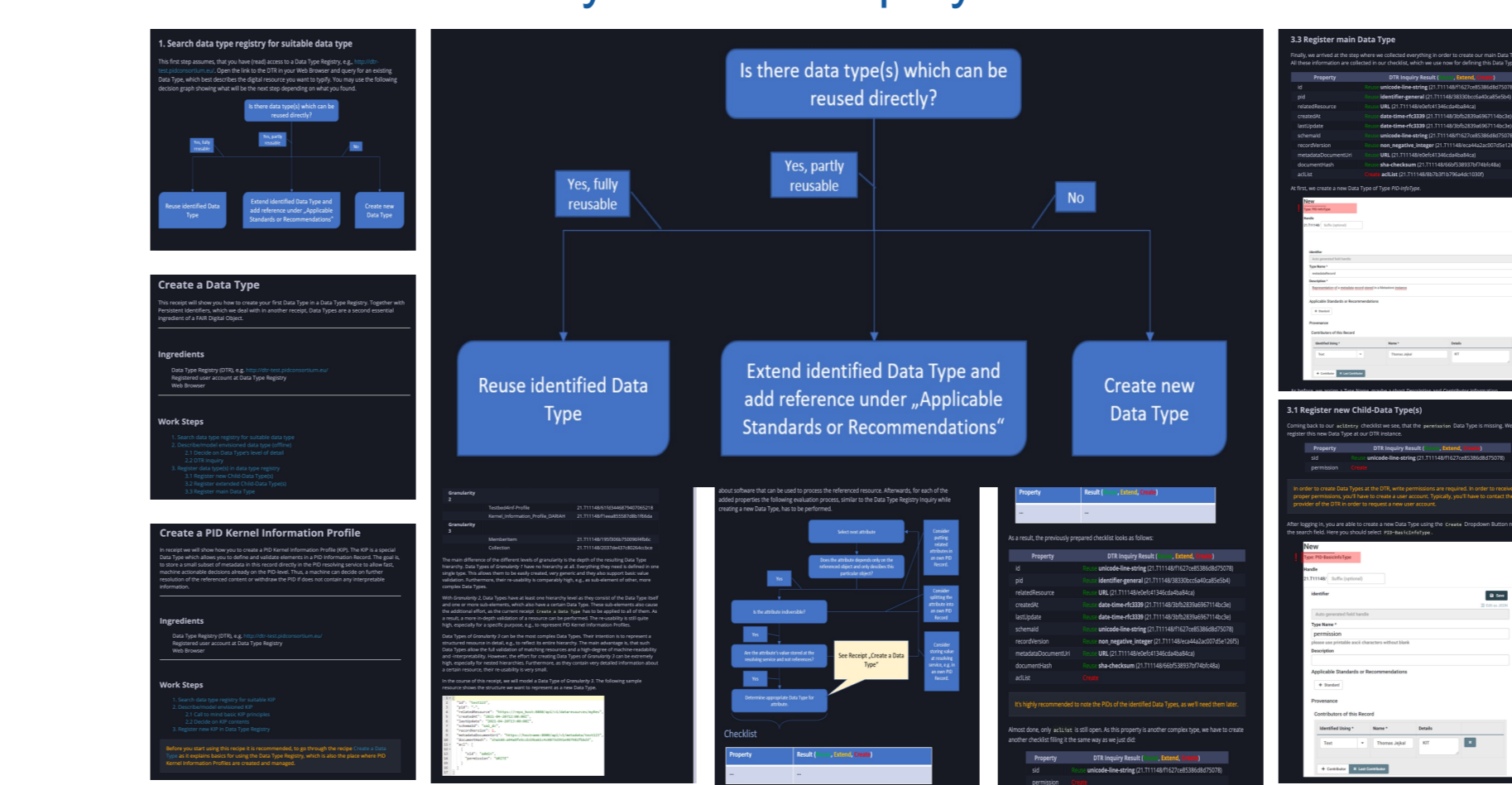
Data Collections Explorer



- Source code: <https://rxp.datamanager.kit.edu>

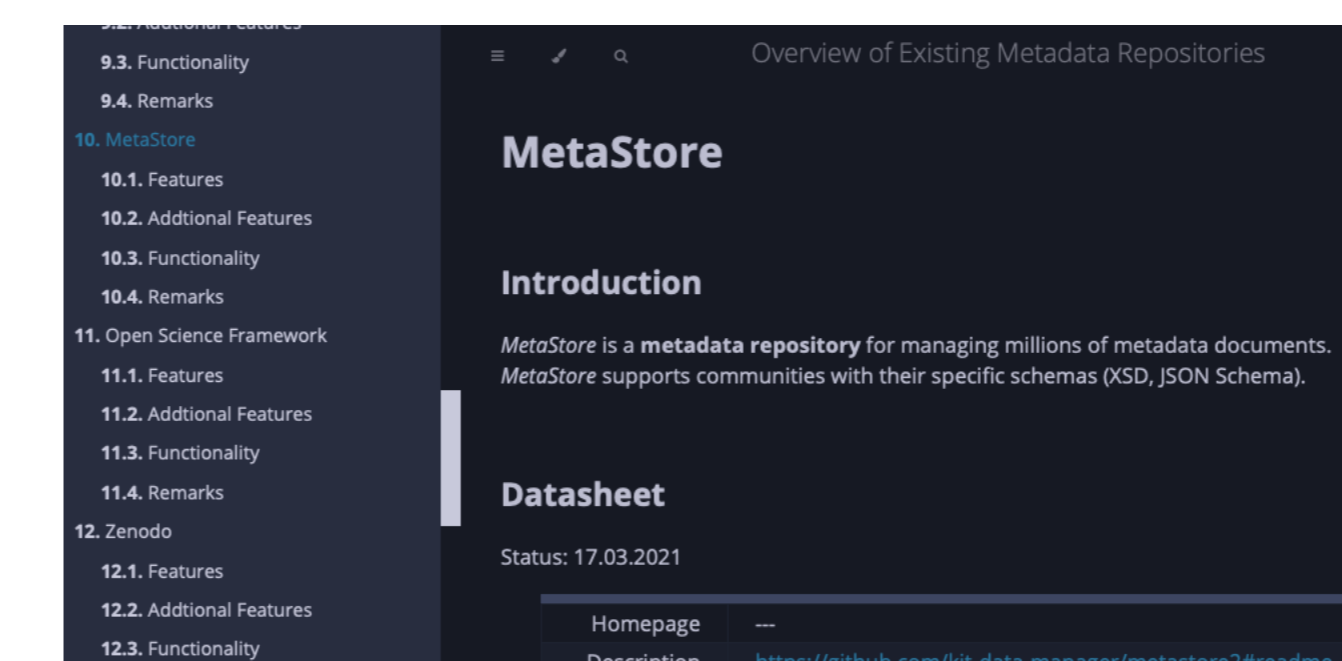
FAIR Digital Object Cookbook

The Cookbook offers step-by-step guidance and good practice for domain data experts integrating the FAIR DO concept. The implications of creating a new data type in terms of effort, reusability and validation are analysed and displayed to the user.



- Uniquely identify resource with implicit characteristics: Effort, Reusability, Validation
- Uniquely identify resource and certain properties: Effort, Reusability, Validation
- Uniquely identify resource and its structure: Effort, Reusability, Validation

Catalog of Repository Platforms



- Source code: <https://kit-data-manager.github.io/repository-platforms/>

Metadata Schemas

Metadata Schema for Scanning Electron Microscopy

Challenges:

- Heterogeneous type of information
- Different instruments
- Different coordinate systems
- Across different length scales
- At different times
- From different regions of the sample
- Under different measuring conditions

Why Metadata Management?

- Find results
- Access data
- Interoperate for exchange and comparison
- Reuse and reproduce

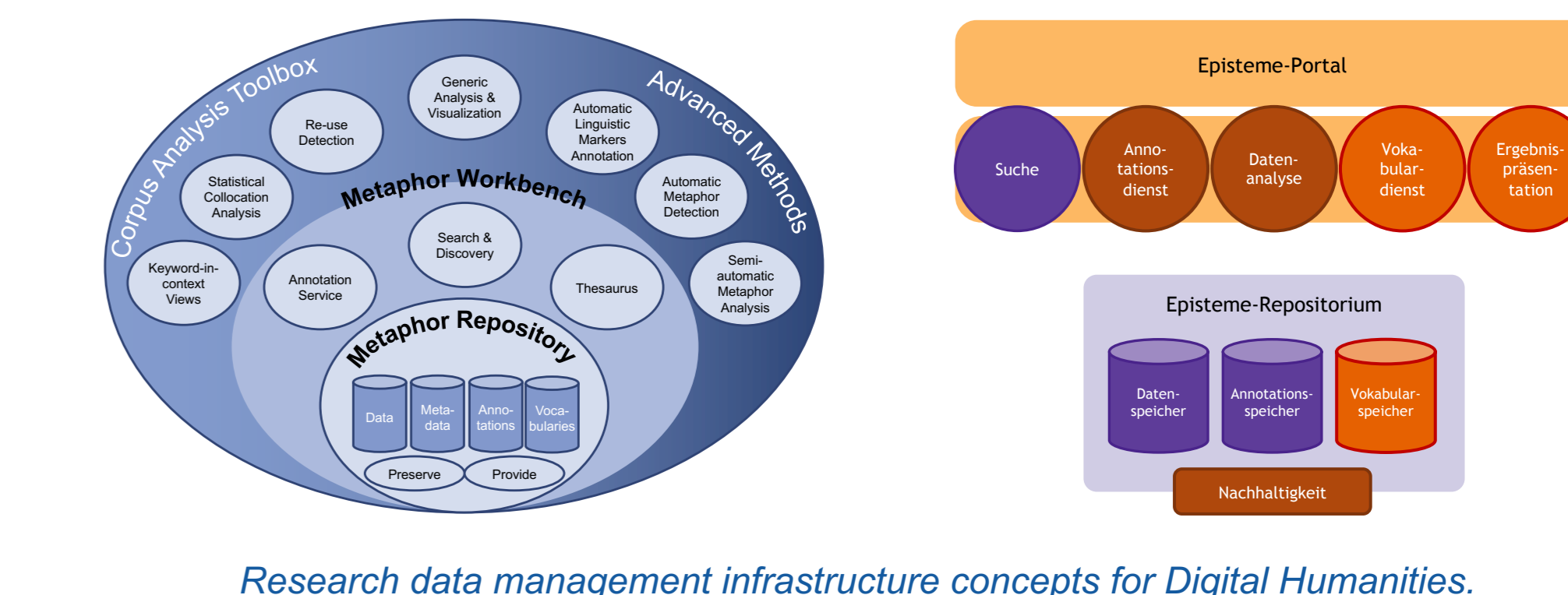
DOI: [10.5445/IR/1000141604](https://doi.org/10.5445/IR/1000141604)

DOI: [10.5445/IR/1000140964](https://doi.org/10.5445/IR/1000140964)

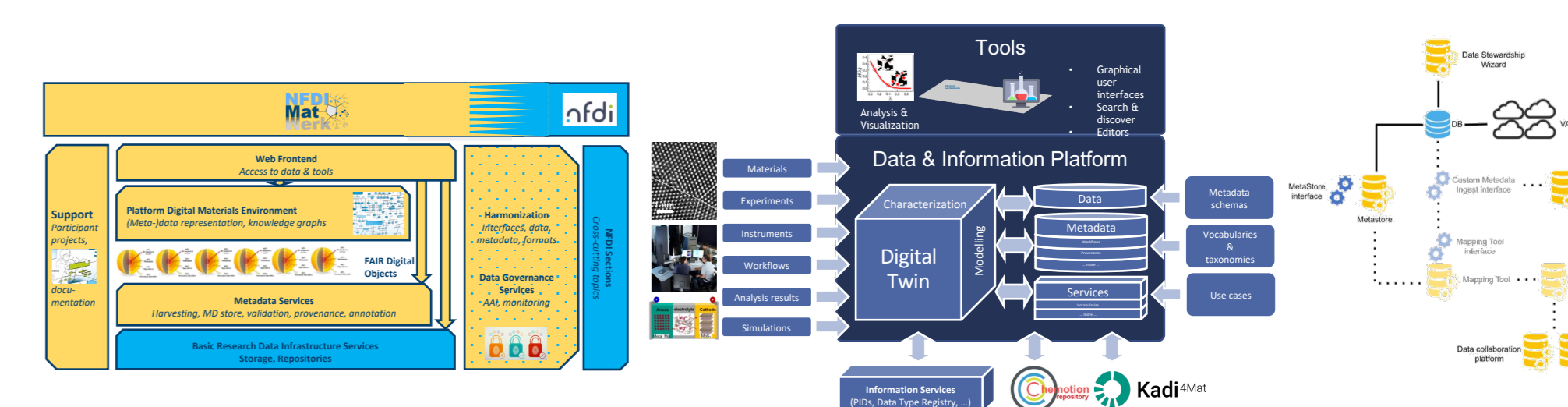
Metadata Schema and FAIR DOs for Machine Learning

The FAIR Digital Object (FAIR DO) concept can be applied to simplify the composition of training data sets for Machine-Learning (ML) applications. For that, a FAIR DO profile is used that has been equipped with a simple metadata schema providing the category information required for supervised classification.

Reference Architectures



Research data management infrastructure concepts for Digital Humanities.



Research data management infrastructure concepts for Materials and Nano Sciences. The service components and tools are the basis for domain-specific high-level services and community-specific tools.