



EUDAT B2FIND : A Cross-Discipline Metadata Service and Discovery Portal

Heinrich Widmann and Hannes Thiemann

DKRZ, Datamanagement, Hamburg, Germany (widmann@dkrz.de)

The European Data Infrastructure (EUDAT) project aims at a pan-European environment that supports a variety of multiple research communities and individuals to manage the rising tide of scientific data by advanced data management technologies.

This led to the establishment of the community-driven Collaborative Data Infrastructure that implements common data services and storage resources to tackle the basic requirements and the specific challenges of international and interdisciplinary research data management.

The metadata service B2FIND plays a central role in this context by providing a simple and user-friendly discovery portal to find research data collections stored in EUDAT data centers or in other repositories. For this we store the diverse metadata collected from heterogeneous sources in a comprehensive joint metadata catalogue and make them searchable in an open data portal.

The implemented metadata ingestion workflow consists of three steps. First the metadata records - provided either by various research communities or via other EUDAT services - are harvested. Afterwards the raw metadata records are converted and mapped to unified key-value dictionaries as specified by the B2FIND schema. The semantic mapping of the non-uniform, community specific metadata to homogenous structured datasets is hereby the most subtle and challenging task. To assure and improve the quality of the metadata this mapping process is accompanied by

- iterative and intense exchange with the community representatives,
- usage of controlled vocabularies and community specific ontologies and
- formal and semantic validation.

Finally the mapped and checked records are uploaded as datasets to the catalogue, which is based on the open source data portal software CKAN. CKAN provides a rich RESTful JSON API and uses SOLR for dataset indexing that enables users to query and search in the catalogue.

The homogenization of the community specific data models and vocabularies enables not only the unique presentation of these datasets as tables of field-value pairs but also the faceted, spatial and temporal search in the B2FIND metadata portal. Furthermore the service provides transparent access to the scientific data objects through the given references and identifiers in the metadata.

B2FIND offers support for new communities interested in publishing their data within EUDAT.

We present here the functionality and the features of the B2FIND service and give an outlook of further developments as interfaces to external libraries and use of Linked Data.