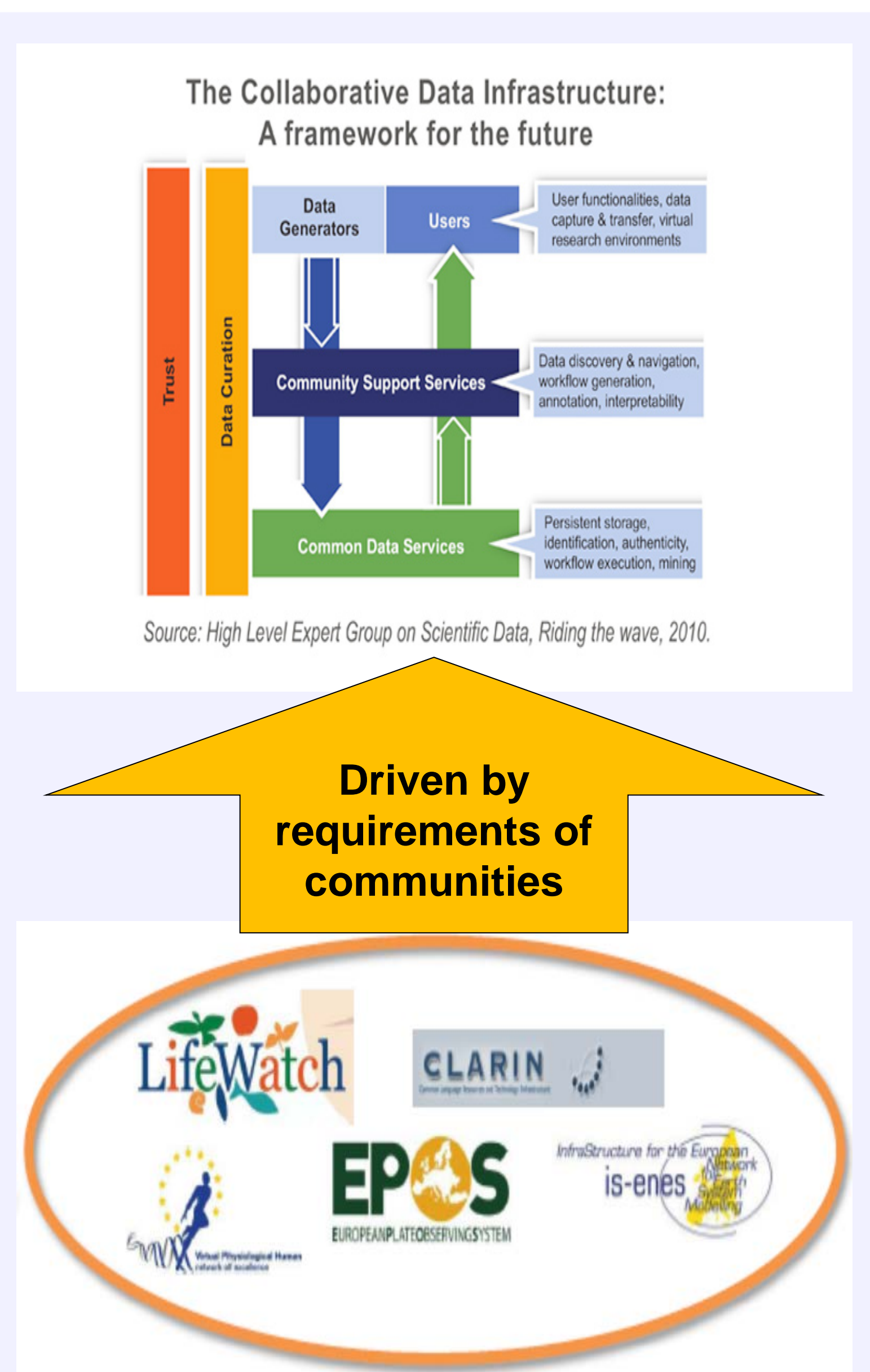


Heinrich Widmann, Hannes Thiemann, and Michael Lautenschlager

EUDAT overview

EUDAT addresses the specific challenges of interdisciplinary and European wide data management of research data. For this a sustainable, cost efficient and high quality **Collaborative Data Infrastructure (CDI)** is built up providing state-of-the-art instruments and services to researchers from all fields on a pan-European level.

The setup and development of the CDI is **driven by** the requirements and needs of the **communities**. For metadata this means that EUDAT has to manage various degrees of complexity across scientific disciplines.



EUDAT services



B2FIND, the EUDAT metadata service, provides a simple and user-friendly way to **find and access** research data stored in EUDAT and external data centers and consists of

- a comprehensive **metadata catalogue** of data objects stored through the B2SHARE and B2SAVE services and harvested from communities which span a large number of datasets from a wide range of disciplines, and
- a web **search portal** allowing researchers to find easily collections of scientific data and access those data through the given references in the metadata.

The B2FIND service has strong interrelation with the other three main services concerning the **storage of data objects**, to which the metadata relate:



B2SHARE is a user-friendly, reliable & trustworthy way for researchers & communities to **store and share** their research data.



B2STAGE is a efficient service to **ship** data between EUDAT storage resources and workspace of **high-performance computing** systems.



B2SAFE is a robust, safe and highly available **replication** service allowing departmental repositories to replicate their research data.

B2FIND metadata ingestion

consists of three workflow steps implemented as independent modules **Harvester**, **Mapper** and **Uploader**. The modular design allows standalone execution and flexible exchange of software components. A controlled workflow framework ensures synchronicity between the provided metadata and the B2FIND catalogue.

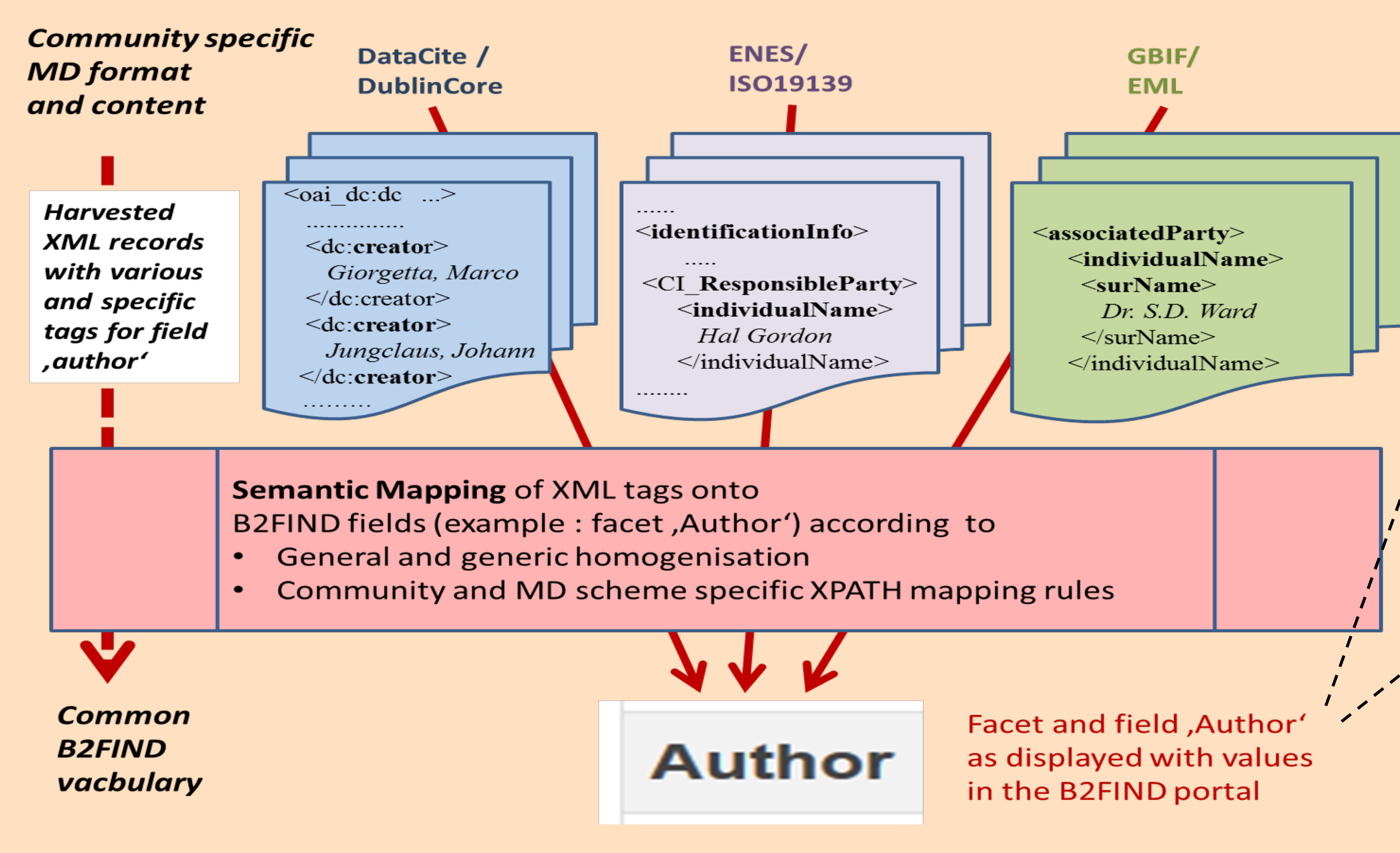
1. OAI-PMH Harvesting

Metadata is harvested from data providers using the standard protocol OAI-PMH. The communities themselves decide which metadata is made available to EUDAT.

2. Semantic Mapping and Homogenisation of Metadata

The most challenging task within B2FIND is the mapping of the metadata compilations realized in various formats and schemes across scientific disciplines. To manage the great diversity we developed an elaborate mapping tool that allows a unique and generic homogenisation of the heterogeneous metadata to a common vocabulary.

We illustrate this in the following figure by the example of translation of appropriate entries in the raw XML records to the **B2FIND facet 'Author'**.

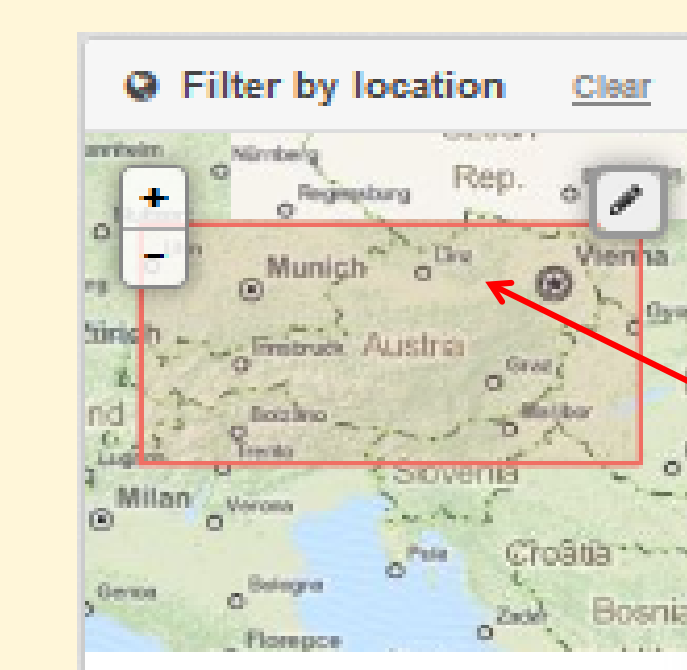


3. Uploading to the B2FIND catalogue and portal

The mapped JSON records (i.e. dictionaries of mapped key-value pairs) are uploaded to the B2FIND portal and keys are SOLR-indexed to enable efficient search.

B2FIND portal → b2find.eudat.eu

is based on the data portal platform CKAN (ckan.org), supports **user-friendly navigation** and offers many powerful **search functionalities**.



Faceted search

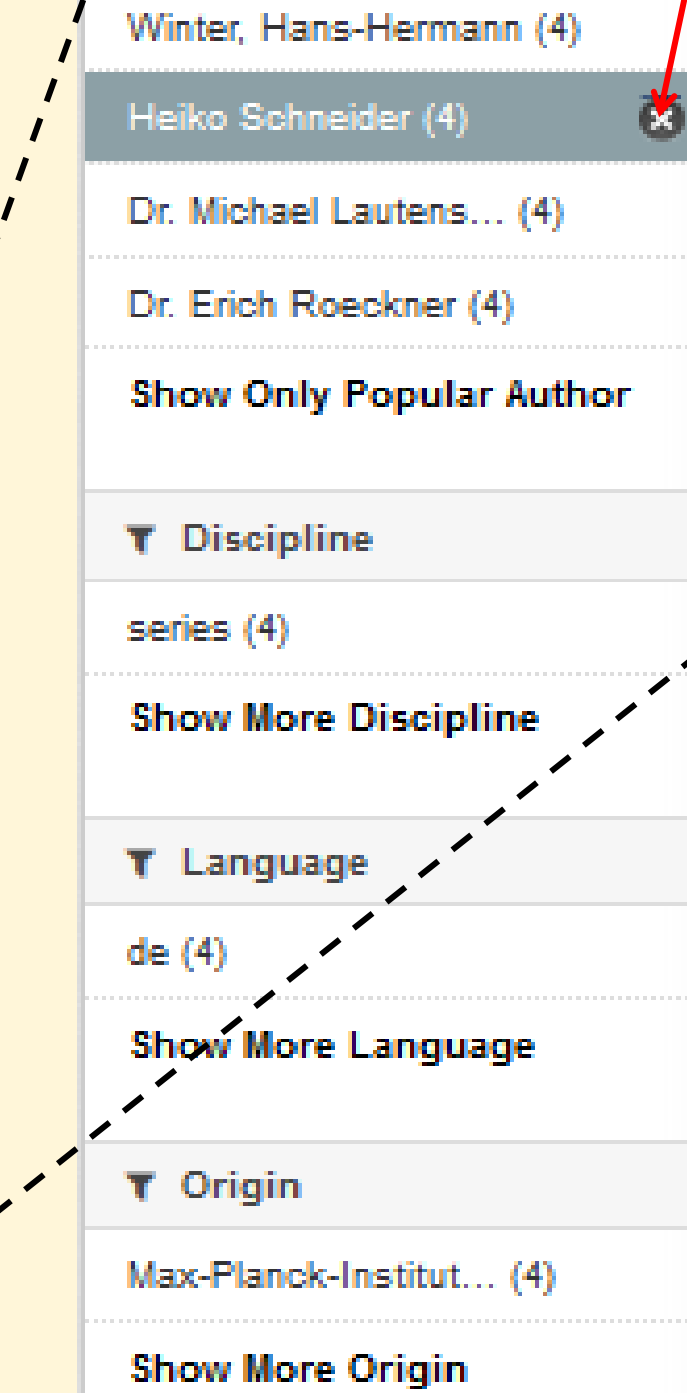
by **selection of values from B2FIND facets** (i.e. searchable fields) or by **spatial or temporal filtering** in the navigation bar:

- **Geospatial selection** of a region from the world map
- **Publication Year**: selection of years dataset is published in
- **Full text search** using the free text field
- **Faceted search**: selection of a value from facets

Search requests can be combined and results in the hit list of datasets.

Dataset view

by opening a dataset **Spatial extent** is displayed and **Title and description** are shown in the upper range. Further **B2FIND fields and their values** are displayed in tabular form.



Field	Value
Source	http://dx.doi.org/DOI:10.1594/WDC/ENES-T03L31_OM-GR1.5L40_A2_1_0H
Author	Winter, Hans-Hermann; Dr. Erich Roeckner; Dr. Michael Lautenschlager; Heiko Schneider
Maintainer	Max-Planck-Institut fuer Meteorologie (M), Model and Data
Discipline	series
Format	WMO format GRidded Binary data
Language	de
MetadataAccess	http://c3grid1.dkrz.de:8080/oai/provider?verb=GetRecords&metadataPrefix=iso&identifier=oai:c3grid.dkrz.de:dkrz.wdc.iso_doi_iso_old2038017
Origin	Max-Planck-Institut fuer Meteorologie (M), Model and Data
PublicationYear	2004

Data access

References allow access to data sources:

- **Source**: link to the data object the metadata refer to
- **MetadataAccess**: link to the harvested metadata XML record

Outlook and contact

Further development will bring more functionalities and features as commenting datasets and search in taxonomies and for temporal coverage.

B2FIND is open to all researchers free of charge. Research communities who want to publish their metadata within EUDAT should contact us through <http://www.eudat.eu/support-request>. EUDAT is also interested in helping emerging communities to set up their own metadata infrastructure.

