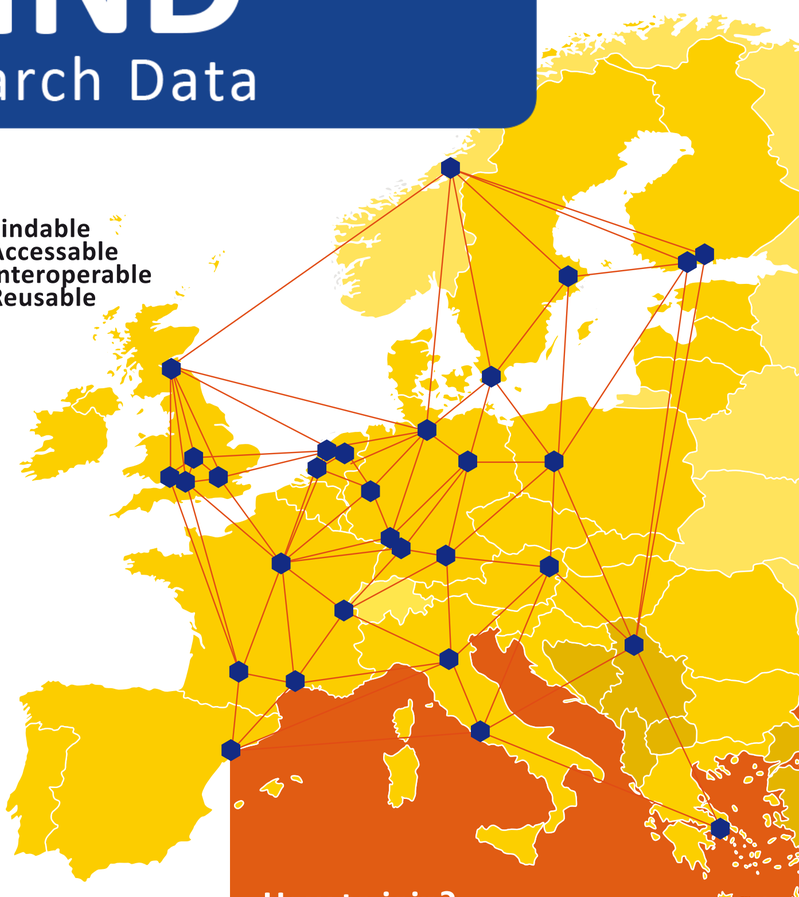
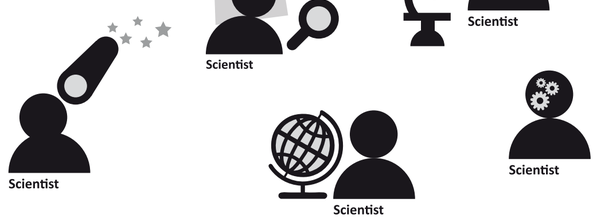
B2FIND
Find Research Data

Making Research Data FAIR

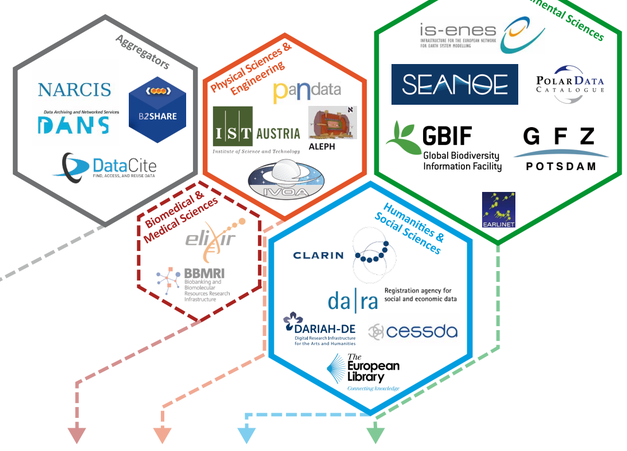
Findable
Accessable
Interoperable
Reusable



Scientist might be one of the **Researcher**

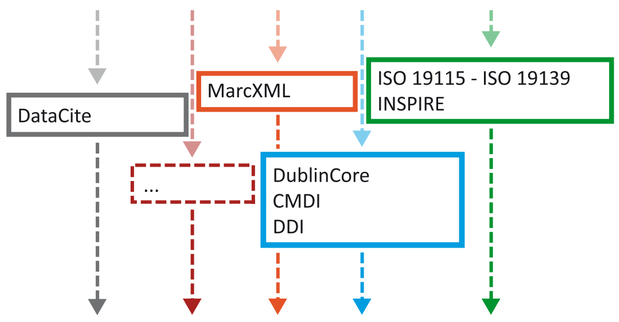


of **B2FIND - Communities**



which do **Data/Metadata generation**

Scientific communities are using different metadata schemas in order to display and utilize their research data. B2FIND already supports several metadata standards, new ones can be adapted if necessary.



Harvesting by B2FIND

Harvesting is the process of automatically fetching remote metadata. Research data stored within EUDAT data centres (via B2SAFE/B2SHARE) are automatically represented in B2FIND if desired. B2FIND preferably uses OAI-PMH as the simplicity of the protocol allows a controlled and easy to manage transfer of metadata, but other APIs are supported as well.

OAI-PMH
Only little information must be provided to enable B2FIND to perform the harvesting process:

- `OAI endpoint`: URL of the OAI provider server on data provider site, which must be open for OAI-PMH read requests
- `OAI mdprefix`: OAI acronym for the metadata schema in which the provided XML records are coded in
- `OAI sets` (optional): it is recommended to group your records in subsets due to a simplification of the controlled harvesting

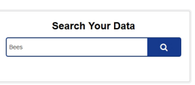
JSON-API
Some data providers offer their metadata encoded as JSON records, which can be retrieved, queried and browsed via a REST API. The API is generally RESTFUL and returns results in JSON, as the API follows the JSONAPI specification.

CSW
Catalog Service for the Web (CSW) is a standard for presenting geospatial records in XML. The catalogue is made up of records that describe geospatial data and services. B2FIND uses a CSW 2.0 implementation to harvest XML records from so called GEO network portals.

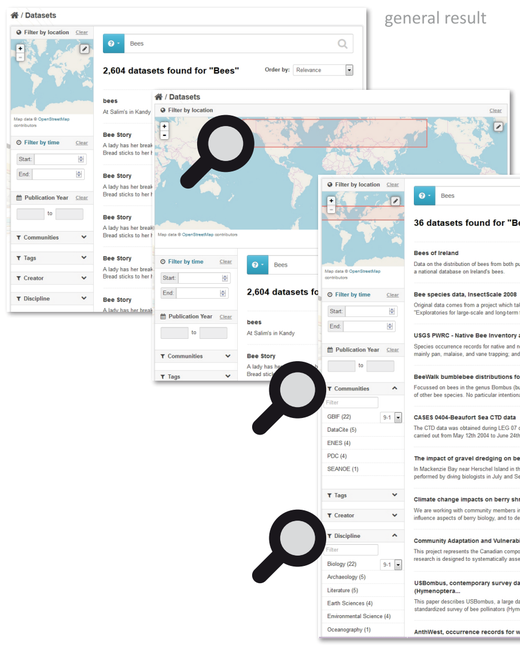
General information
eudat.eu/services/b2find
http://b2find.eudat.eu

Contact
martens@dkrz.de
widmann@dkrz.de

Posterdesign: Daniela Heigl & Claudia Martens



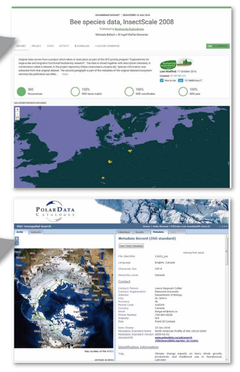
general search term



general result

narrow down by geospatial search

narrow down by facets, e.g.:
- Communities
- Disciplines



landing pages for data access

How to join?

For making your resarch data openly accessible in B2FIND this few requirements are needed:

- An API to make metadata available for harvesting must be set up. OAI-PMH is preferred, but other APIs are equally supported.
- Metadata should be delivered in some standaradized format and schema. Various metadata schemas are promoted, new ones can be adapted.
- Community specifications, structures and data models have to be communicated.

All other refinements and elaboration of the mapping can be iteratively discussed later on.

Guidelines for further information:
http://b2find.eudat.eu/guidelines/introduction.html

B2FIND

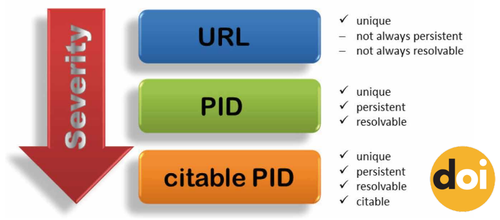
is a powerful and user friendly discovery portal that enables scientists from all over the world to find and access data collections from heterogeneous sources via a web interface. It offers metadata of research data that are stored in EUDAT data centres as well as metadata that are steadily harvested from community specific repositories.
B2FIND covers a wide range of research areas: Climate Research to Social Sciences, Biodiversity to Linguistics or Archaeology to Seismology. Searching through cross-domain sources and finding dispersed data improves interoperability and reusability of research data and hence fosters open science based on FAIR principles.

Providing research metadata on the web

B2FIND provides an interdisciplinary and cross-community search with geospatial and temporal search functionalities, whereas results from a free text search may be filtered and narrowed down by using the facets.

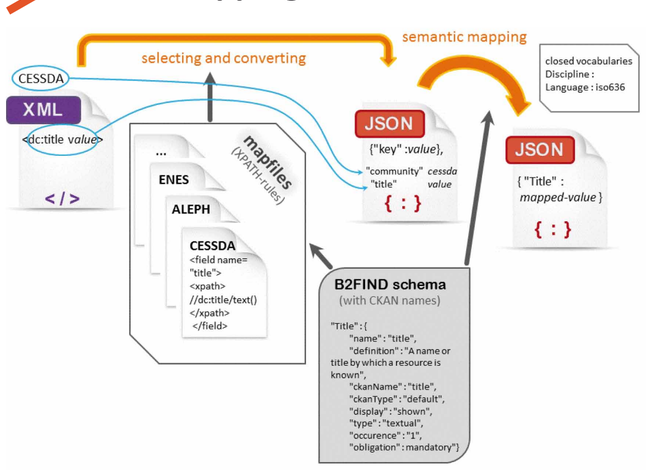
Harvested and mapped metadata records of research datasets are uploaded in a CKAN database and searchable via a web interface. Metadata are made publicly available and openly accessible under a CC BY-SA 4 license - even though access to the datasets itself may be restricted.

Identifier



Knowing that the amount of digital research data continuously increases both in volume and in numbers of digital objects, B2FIND sets great store by reliable and persistent identification of digital objects.
B2FIND's metadata schema separates three elements for a unique resource reference: `Source`, `DOI` and `PID`. At least one identifier is mandatory, for best citability of datasets usage of `DOIs` is recommended.

Mapping onto B2FIND schema



B2FIND's mapping process is twofold:
Initially metadata elements from harvested XML records are selected and converted into JSON files with community specific defined 'key' 'value' pairs onto the B2FIND schema.
During the semantic mapping an iterative process extracts further relevant information, adds new content and maps external content.

B2FIND SCHEMA

general information
`<title>`
`<description>`
`<tags>`

access data
`<source>`
`<doi>`
`<pid>`
`<checksum>`
`<rights>`

additional information
`<contact>`
`<metadaAccess>`

provenance data
`<community>`
`<discipline>`
`<creator>`
`<publisher>`
`<publicationYear>`

formal data
`<language>`
`<temporalCoverage>`
`<spatialCoverage>`
`<format>`
`<resourceType>`

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.

DKRZ
DEUTSCHES KLIMARECHENZENTRUM