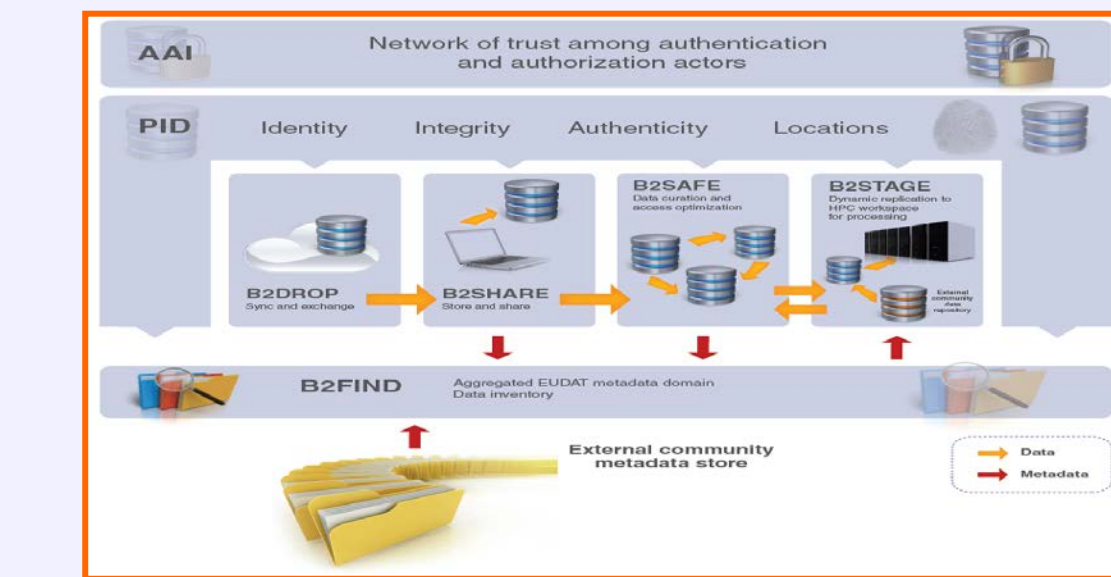


EUDAT aims at a pan-European environment and built up a sustainable and high quality **Collaborative Data Infrastructure (CDI)** that

- is **driven by** the requirements and needs of a variety of multiple research **communities** and individuals,
- addresses the specific challenges of **interdisciplinary research data management** and
- provides advanced data management **services to manage the rising tide of data and various degrees of complexity**



Development of the CDI is driven by requirements of communities



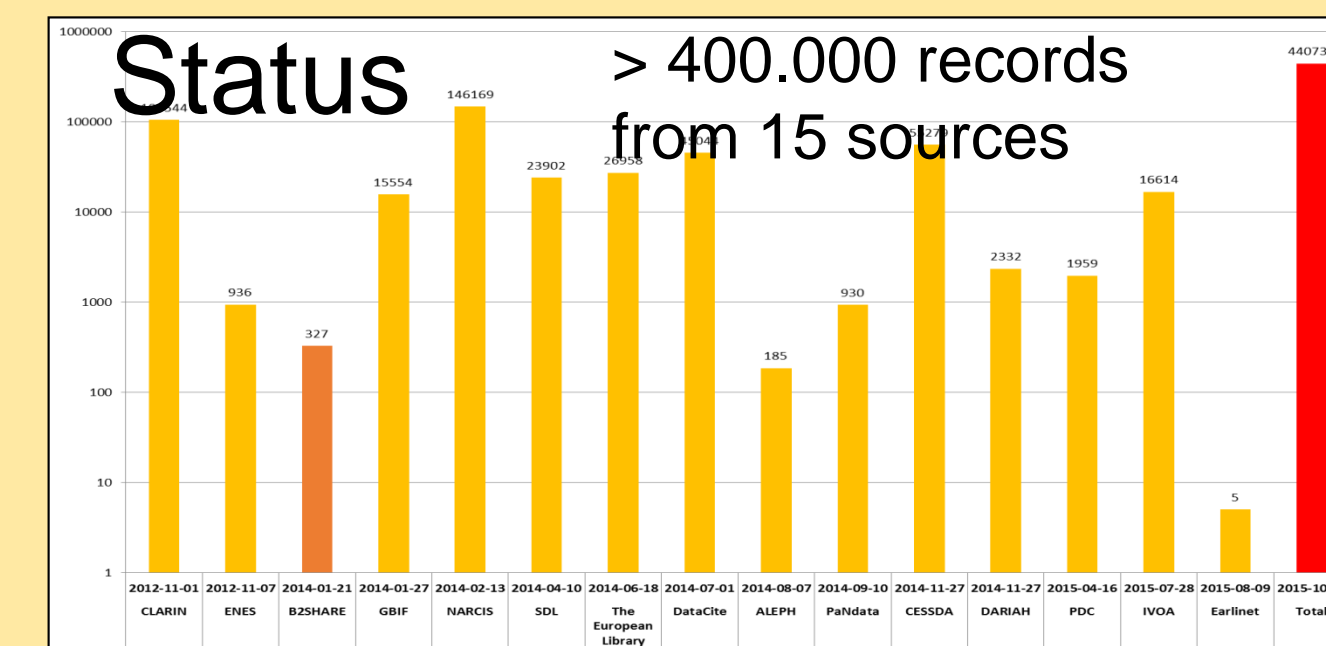
EUDAT CDI and B2 Services

B2FIND is the EUDAT **metadata (MD) service** that

- provides a user-friendly **discovery portal** on data collections from a wide range of research communities
- is based on a joint **MD catalogue** of research data stored in EUDAT data centres and other repositories

Motivation or why should you **publish your MD** in EUDAT ?

- Data get **global search- and accessible** in a **cross-disciplinary** scope
- Improved **interoperability** and re-use of research data
- Feedback** and annotation on research output
- Benefit from **quality assurance and added value** of your metadata



Ingestion Workflow

MD Generation should be in close proximity to the data production to benefit from early quality control



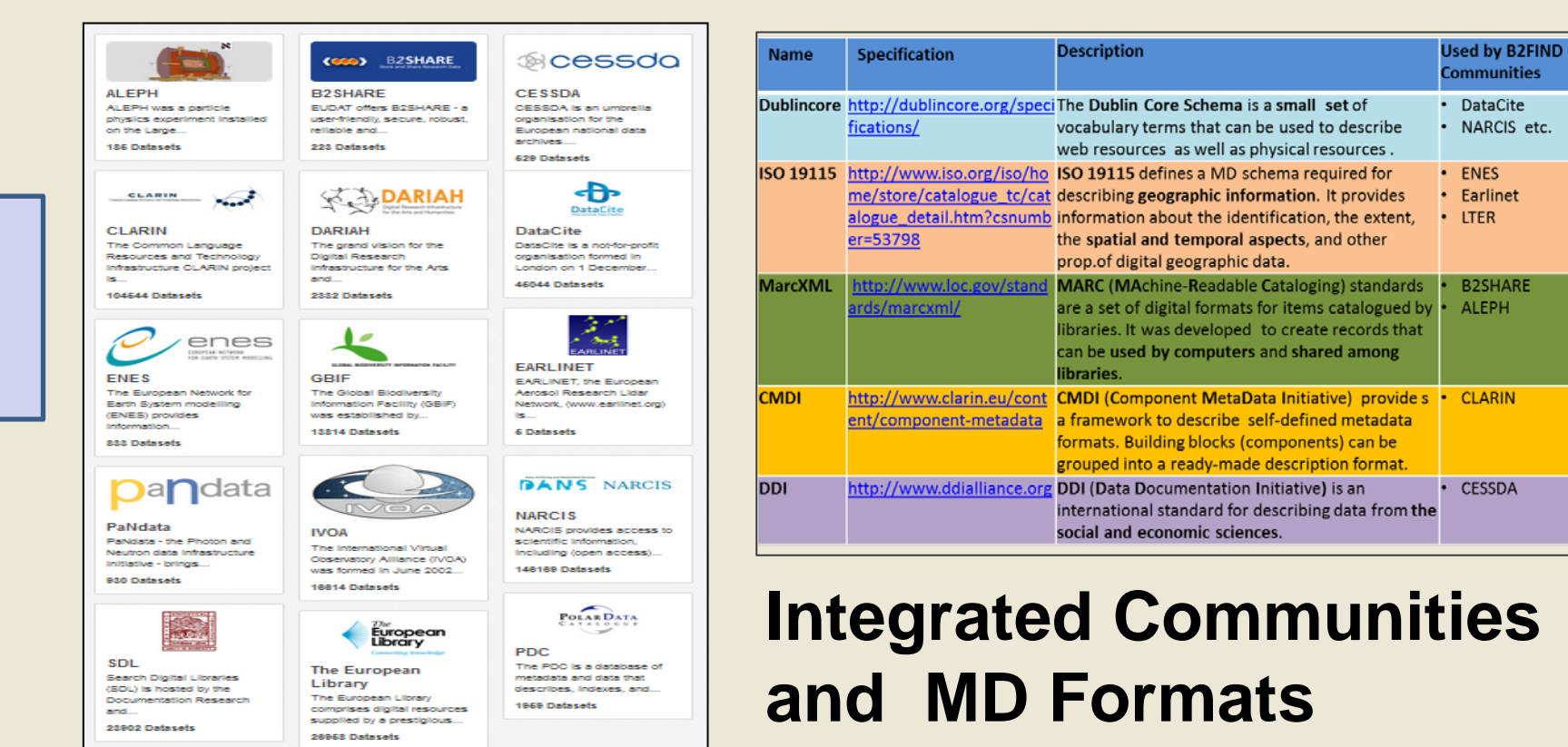
MD Harvesting is performed regularly and incrementally from OAI endpoints of various sources

MD Semantic Mapping, i.e. community specific MD homogenisation to B2FIND schema :

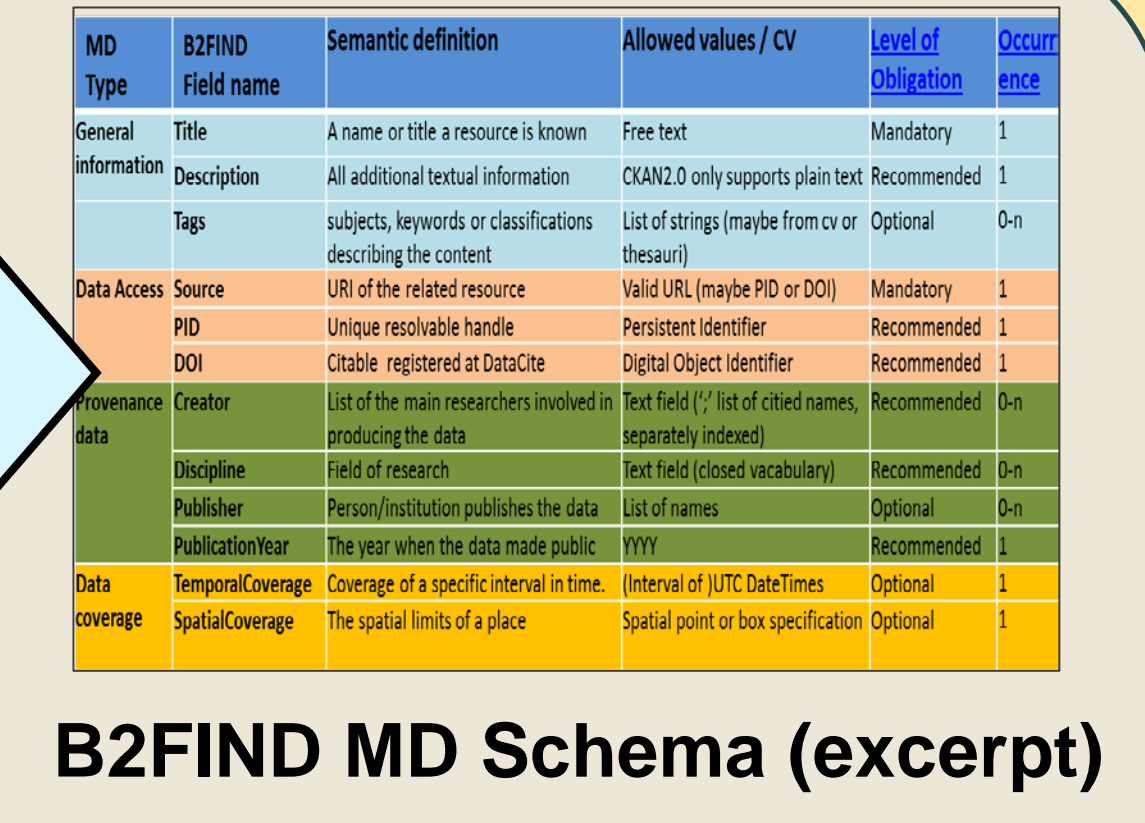
- Select entries from XML records by MD format specific XPATH rules
- Parse and analyse values and map to key-value pairs (JSON), using (comm. specific) controlled vocab's

MD Validation by checking each key-value pair for coverage, consistency and validity and technical checks, e.g. date-time vs. UTC format. spatial coverage by geonames.org and consistency of lat/lon coord.

MD Uploading and Indexing of the mapped and checked JSON records as datasets to the catalogue

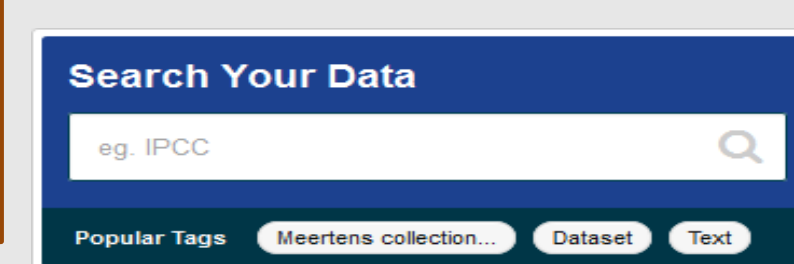


Homogenisation



Search and Access

Search and Browse datasets via Keyword searches

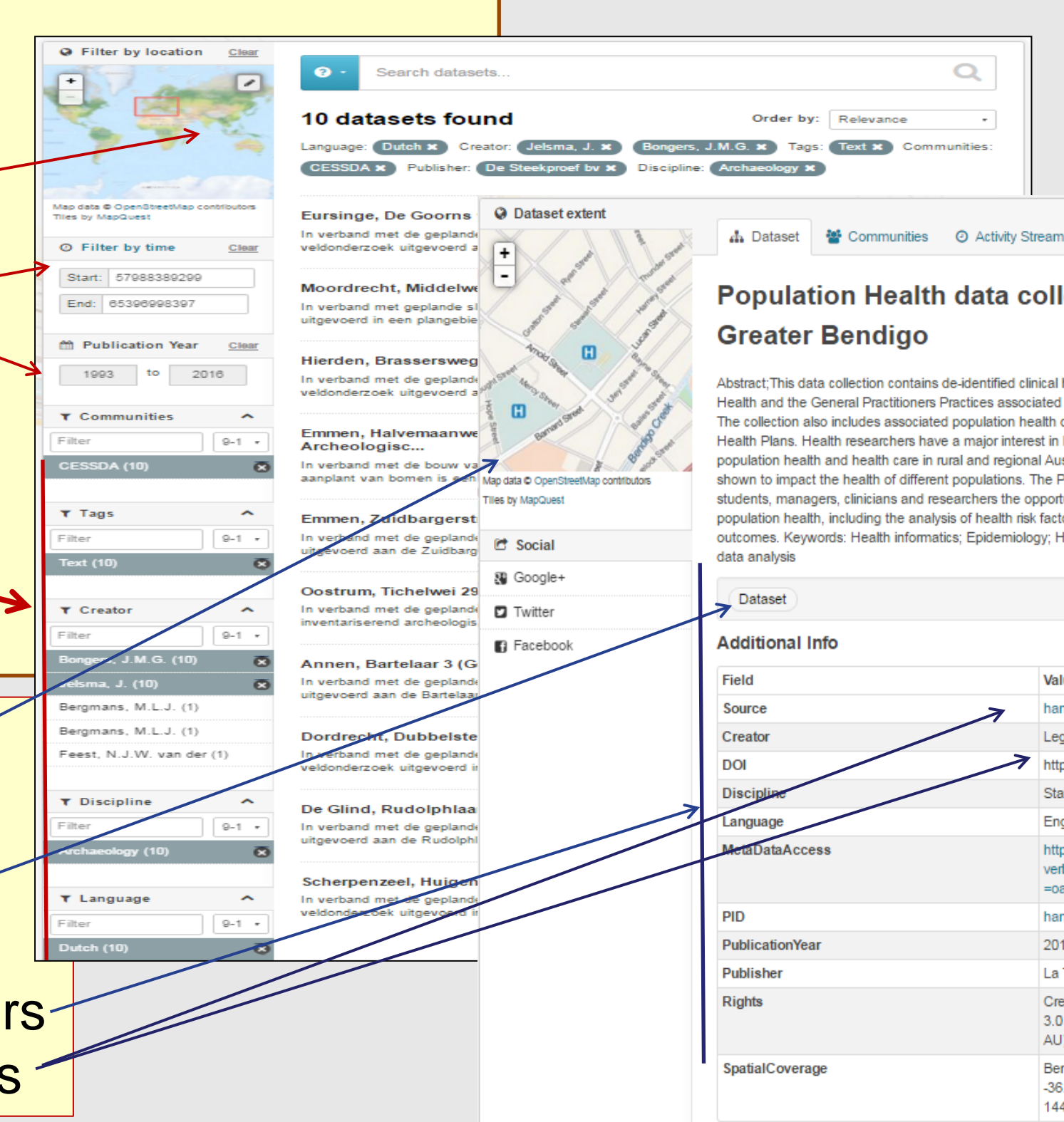


Command line interface using the CKAN API is possible via the script
\$ searchB2FIND.py
[-c comm] facet:value

Results displayed in easy to read format and listed in order of relevance

Facetted Search for

- Free text
- Geo spatial**
- Temporal coverage**
- Publication year**
- Textual facets as
 - Tags**
 - Creator**
 - Discipline**
 - Publisher**



Dataset view displays

- Spatial extent
- Temporal extent
- Selected tags
- Table of field-value pairs
- Links to data resources

Data access is provided by

- Source** leads to (landing page of) data collection
- DOI and PID** refer to other (landing pages) of the data
- MetadataAccess** displays the originally harvested metadata

EUDAT B2FIND MD Catalogue
based on the open source software CKAN (<http://ckan.org/>)

- provides a rich RESTful JSON API
- uses SOLR for dataset indexing
- extendable by adaptable modules

Distribution and Dissemination (work in progress)

Provide B2FIND MD via OAI-PMH to enable other catalogues to harvest the B2FIND catalogue

Annotation of datasets (B2NOTE) allows user to annotate a dataset with a concept.

Use Linked data and the **semantic web** to enrich the MD with additional information

Add interfaces to external libraries as the **Search/Retrieval via URL (SRU) protocol**

External Harvester and Catalogues



Contact, support and documentation is provided at

- Information about B2FIND <http://eudat.eu/services/b2find>
- The B2FIND portal <http://b2find.eudat.eu>
- For support go to <http://eudat.eu/support-request>

Upcoming enhancements comprise beside addressing and integrating **more communities**

- the improvement of **portal functionality as discussed above** :
- Customisation** (templates and extendable facets for specific community needs,)
- Improve **quality of metadata** (Enhance mapping, validation and feedback mechanism)