EUDAT-B2FIND

A FAIR and Interdisciplinary Discovery Portal for Research Data

Heinrich Widmann, DKRZ
Claudia Martens, DKRZ
Open Science Days,
Berlin, 17 October 2017

www.eudat.eu

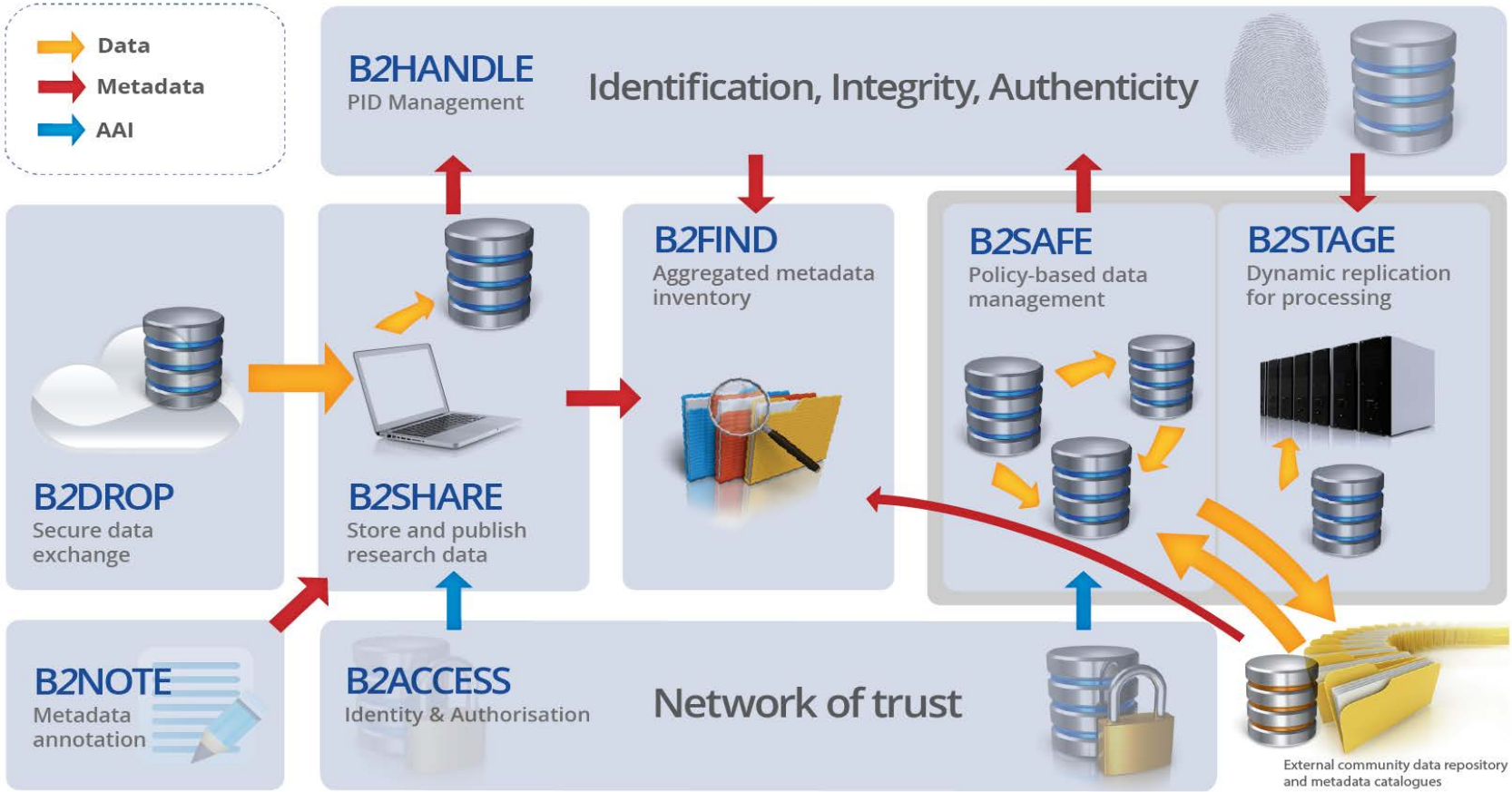
EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

Outline

- **What** is **EUDAT-B2FIND** ?
- **Which guidelines** are followed ?
- **How** we tackle the issue **'diversity vs. interoperability'** ?
- **How** is the service **implemented** ?
- **Which data and domains** does B2FIND **cover** ?
- **How** are the metadata **collected** ?
- **How** can you **search** the Discovery Portal?

What is EUDAT ?

- The project EUDAT (see http://eudat.eu) established a Common Data Infrastructe providing several generic services for inter-disciplinary data management, driven by the communities



What is EUDAT-B2FIND ?

- B2FIND is the metadata service of EUDAT consisting of
  - An interdisciplinary metadata catalogue that spans a large number heterogeneous datasets
    - harvested from various research communities and stored through the EUDAT service B2SHARE
    - covering a wide range of highly diverse disciplines

    ➢ mapped to a common and unified schema
  - An open search portal allowing researchers
    - to find collections of scientific resources in a wide spread and cross-domain search space
    - to access those resources through the given references in the metadata

Which guidelines are followed ?
The FAIR principles → B2FIND approach

- **F**indability →**Discovery Portal** with powerful search features based on rich metadata catalogue
- **A**ccessibility → **Persistent Identifiers** for unique resolvability of data objects
- **I**nteroperability → **Interdisciplinary Catalogue** based on **Common Standards, Vocabs and Schema**
- **R**euseability → **Licenses, Provenance and Domain-relevant Information** is provided

How can data interoperability improved ?

Levels of Interoperability

Heterogeneity ← → Homogeneity

Disciplines

Research Communities (Data Provider)

Cross-disciplinary Data Repositories (e.g. B2SHARE or DataCite)

Service Provider (e.g. EUDAT-B2FIND)

Mandatory

Optional / Recommended

Discipline specific

Schema A

Schema B

Schema C

Collect and extract MD

Harvest MD

DataCite

B2FIND harvest and mapping

B2FIND

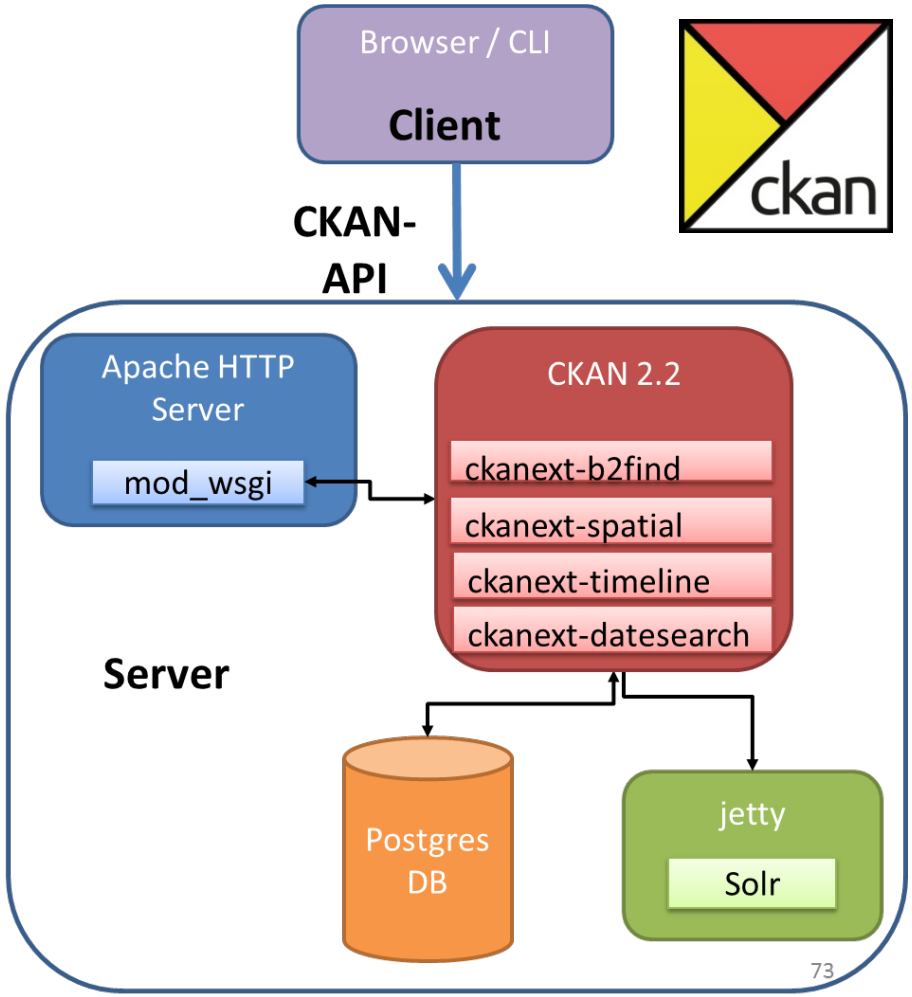
Harvest and Mapping of select MD entries

Extented templates for Domain specific fields

01010
10101
01010

How is the service implemented ?

- uses standard APIs (as OAI-PMH) to harvest from various data providers
- performs a adaptable mapping of the domain specific metadata onto the unified B2FIND schema
- is based on CKAN (see ckan.org), an open source repository and portal software
  - CKAN comes with a SOLR indexer, a Postgres DB and a power and RESTful API
  - Extensible by open available and own developed CKAN extensions



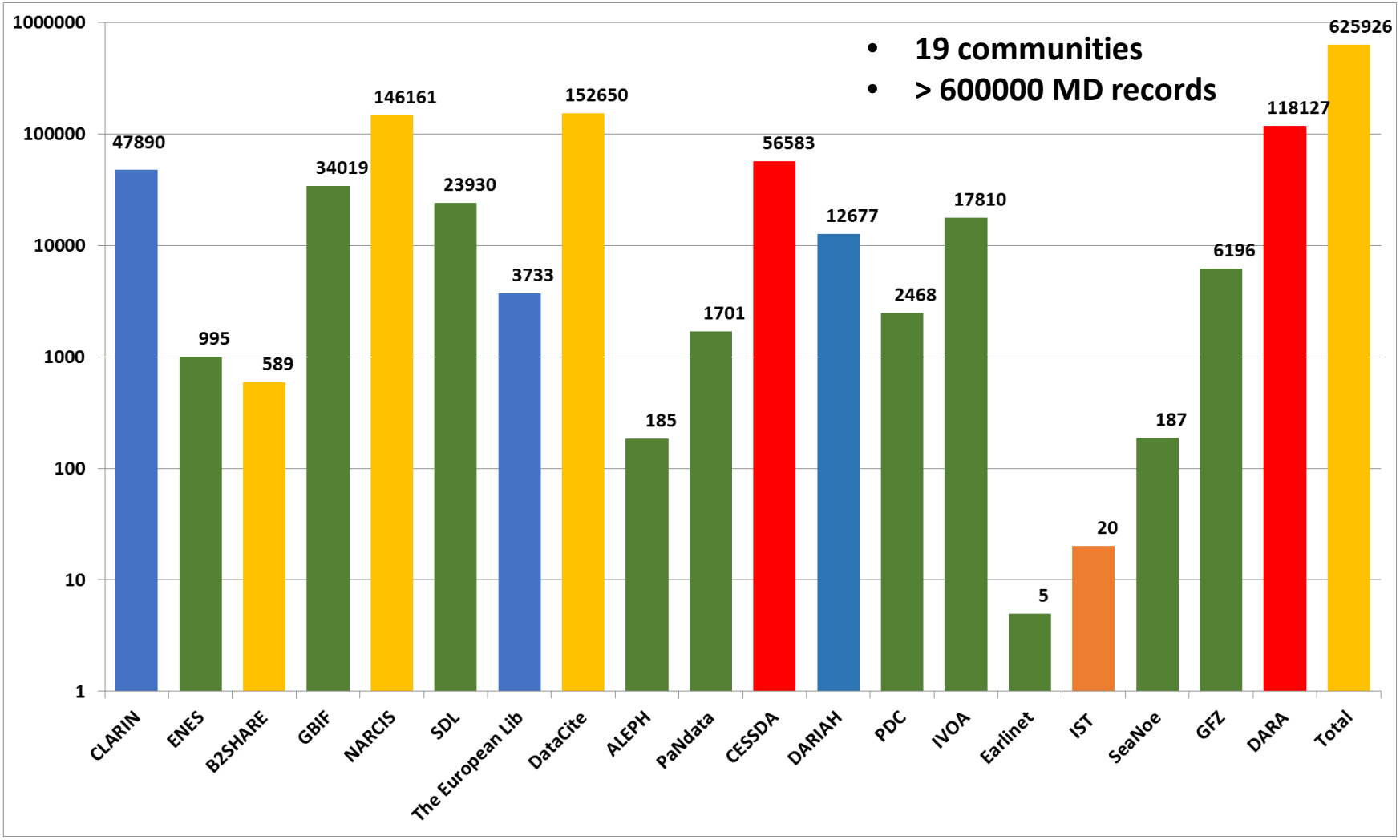
B2FIND MD Schema (extract) based on DataCite 3.0

Metadata Type	B2FIND Field name	Allowed values	Semantic definition	Level of Obligation	Occurenc	
General information	Title	Free text (unicode)	A name or title a resource is known	Mandatory	1	
	Description	Free text	Additional info	Recommended	0-1	
Data Access	Source	Valid URL or URN	Unique link to data resource	Mandatory (1)	0-1	1-3
	PID	Persistent Identifier	+ persistent and resolvable		0-1	
	DOI	Digital Object Identifier	+ citable		0-1	
Provenance data	Creator	';'-sep. list of names	Main researchers involved in data prod.	Recommended	0-n	
	Discipline	List of values from CV	Field of research (Controlled Vocab)	Recommended	0-n	
	Publication Year	YYYY	The year data are published	Recommended	1	
Formal data	Temporal Coverage	Interval of 2 DTimes [Begin, End]	The temporal limits of a date-time	Optional	1-n	
	Spatial Coverage	Spatial box or point [[minlat,minlon…]]	The spatial limits of a place.	Optional	1-n	

Which data and domains does B2FIND covers ?

Humanities
Social Sciences
Natural Sciences
Cross Discipline

Ingestion Status of the MD Catalogue



How can you search the Discovery Portal?

B2FIND Discovery Portal

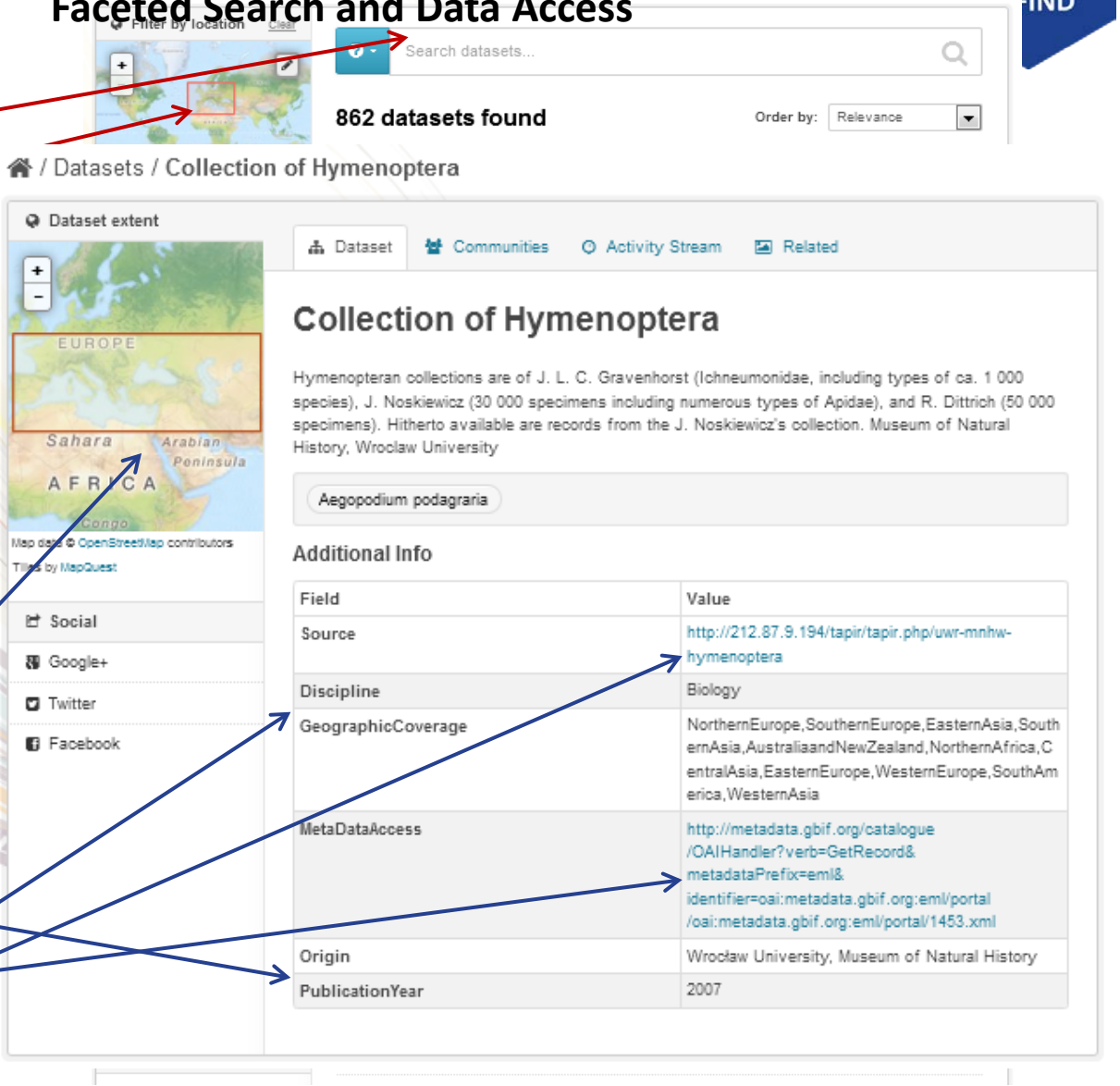
Faceted Search and Data Access

B2FIND provides 'faceted' search for

- Free text
- Geo spatial
- Temporal coverage
- Publication year
- Textual facets as
  - Tags
  - Creator
  - Discipline etc.

Dataset view provides display of metadata :

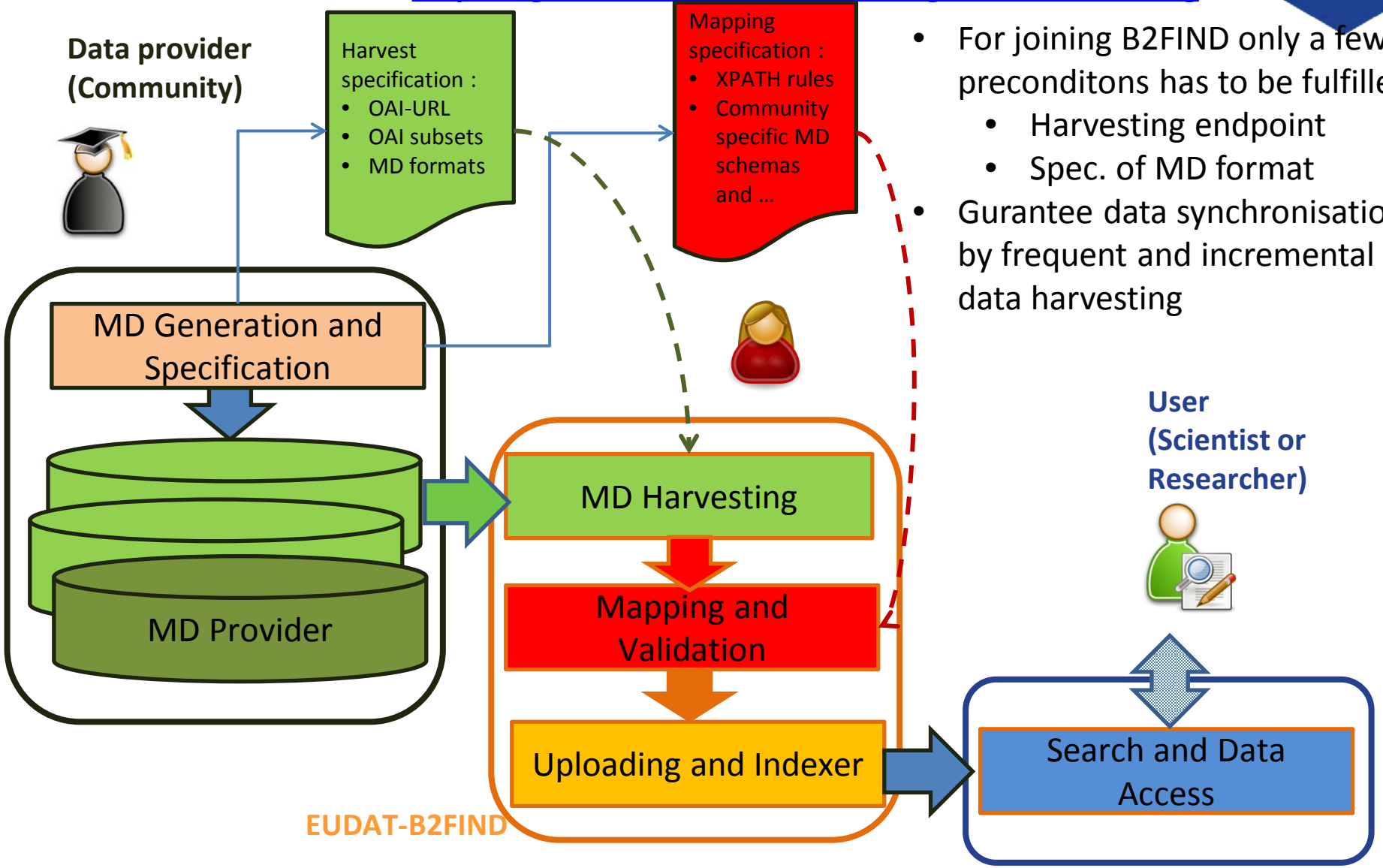
- Spatial extent
- Table of field-value pairs
- Links to data resources



How are the metadata collected ?

B2FIND Ingestion Workflow

→ https://github.com/EUDAT-Training/B2FIND-Training



- For joining B2FIND only a few preconditons has to be fulfilled
  - Harvesting endpoint
  - Spec. of MD format
- Gurantee data synchronisation by frequent and incremental data harvesting

Summary

EUDAT-B2FIND

- established an operative service based on agreed standards and guidelines as the FAIR principles
- provides a discovery portal with powerful faceted search functionalities
- is based on a joint catalogue of research data , combining many heterogeneous and cross-discipline sources
- achieves improved interoperability by homogenisation to a common metadata schema
- follows a low barrier approach for data providers to expose metadata to B2FIND

Thank you for your attention !

Links :

- Info and Docs : http://eudat.eu/b2find
- Guidelines for data providers : http://b2find.eudat.eu/guidelines
- B2FIND Portal : http://b2find.eudat.eu
- Training : https://github.com/EUDAT-Training/B2FIND-Training
- Support : www.eudat.eu/support-request
- Contact : widmann@dkrz.de , martens@dkrz.de