



EUDAT-B2FIND

A FAIR and Interdisciplinary Discovery Portal for Research Data

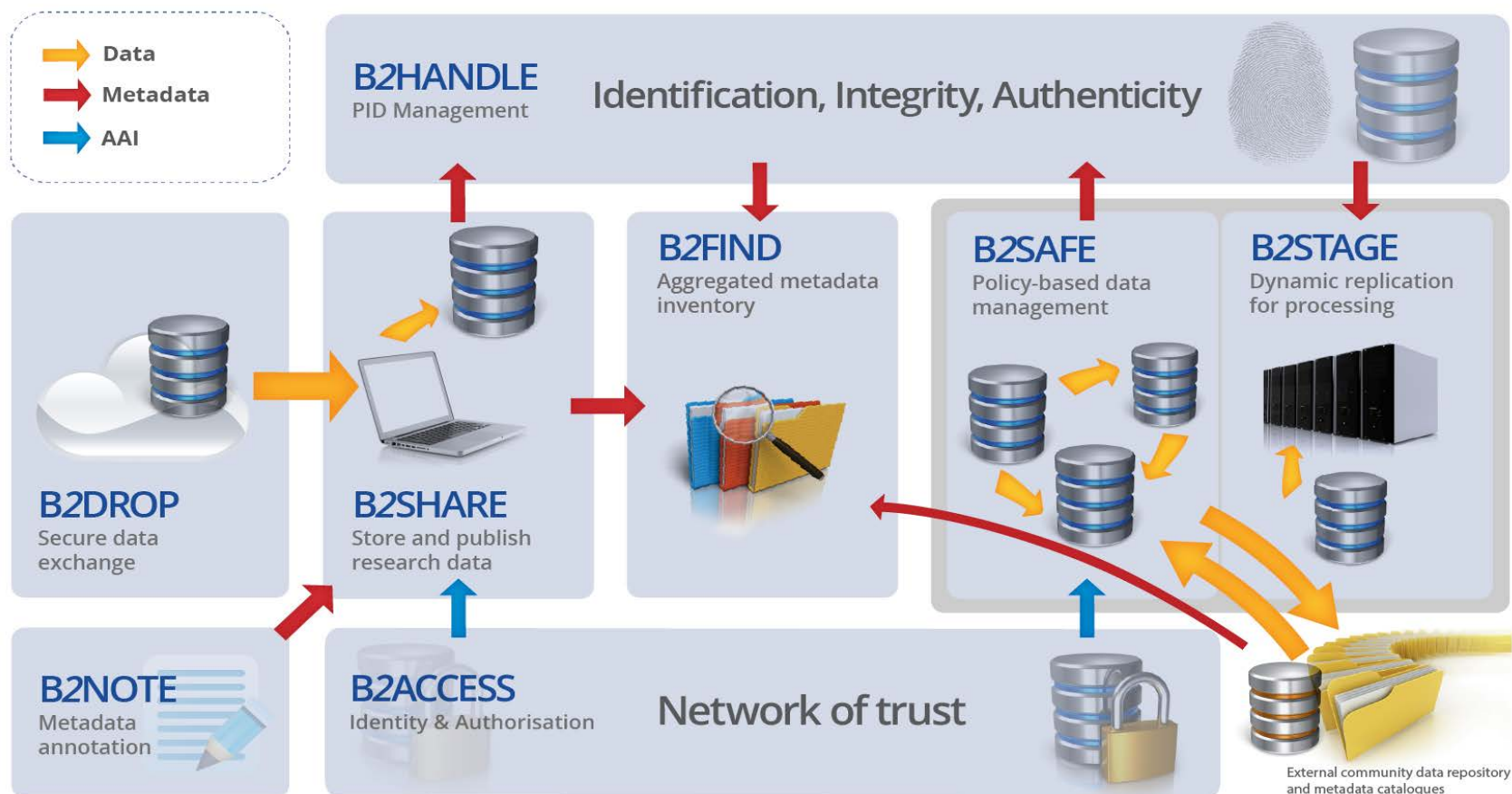
Heinrich Widmann, DKRZ
Claudia Martens, DKRZ
Open Science Days,
Berlin, 17 October 2017



www.eudat.eu

- **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?**

- The project EUDAT (see <http://eudat.eu>) established a Common Data Infrastructure providing several generic services for interdisciplinary 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



- **Findability → Discovery Portal** with powerful search features based on rich metadata catalogue
- **Accessibility → Persistent Identifiers** for unique resolvability of data objects
- **Interoperability → Interdisciplinary Catalogue** based on **Common Standards, Vocab and Schema**
- **Reuseability → Licenses, Provenance and Domain-relevant Information** is provided

How can data interoperability improved ?

Levels of Interoperability

Heterogeneity

Homogeneity

Research Communities
(Data Provider)

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

Service Provider
(e.g. EUDAT-B2FIND)



Disciplines



Optional / Recommended

Discipline specific

01010
10101
01010

BASIC PHYSICS OF NUCLEAR MEDICINE

01010
10101
01010
01010
01010

Schema A

Schema B

Schema C

Mandatory

Collect and extract MD

Harvest MD

DataCite

B2FIND harvest and mapping

Harvest and Mapping of select MD entries

Extended templates for Domain specific fields

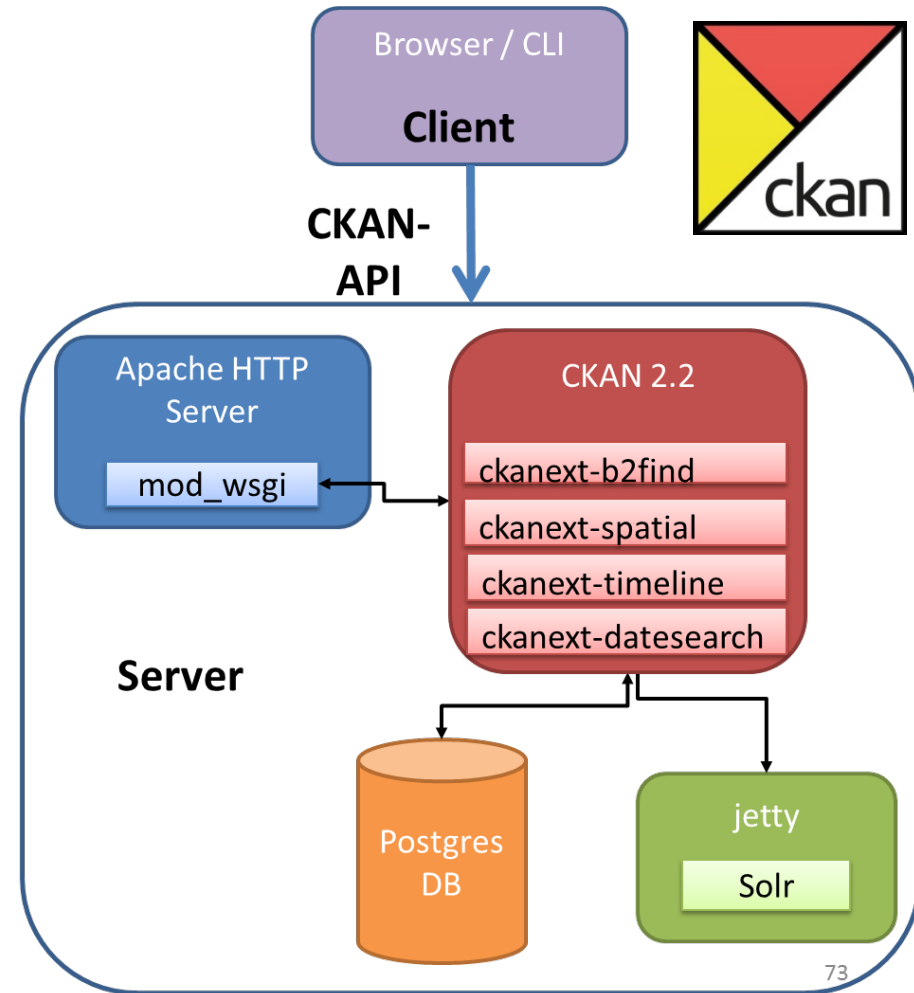
B2FIND



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

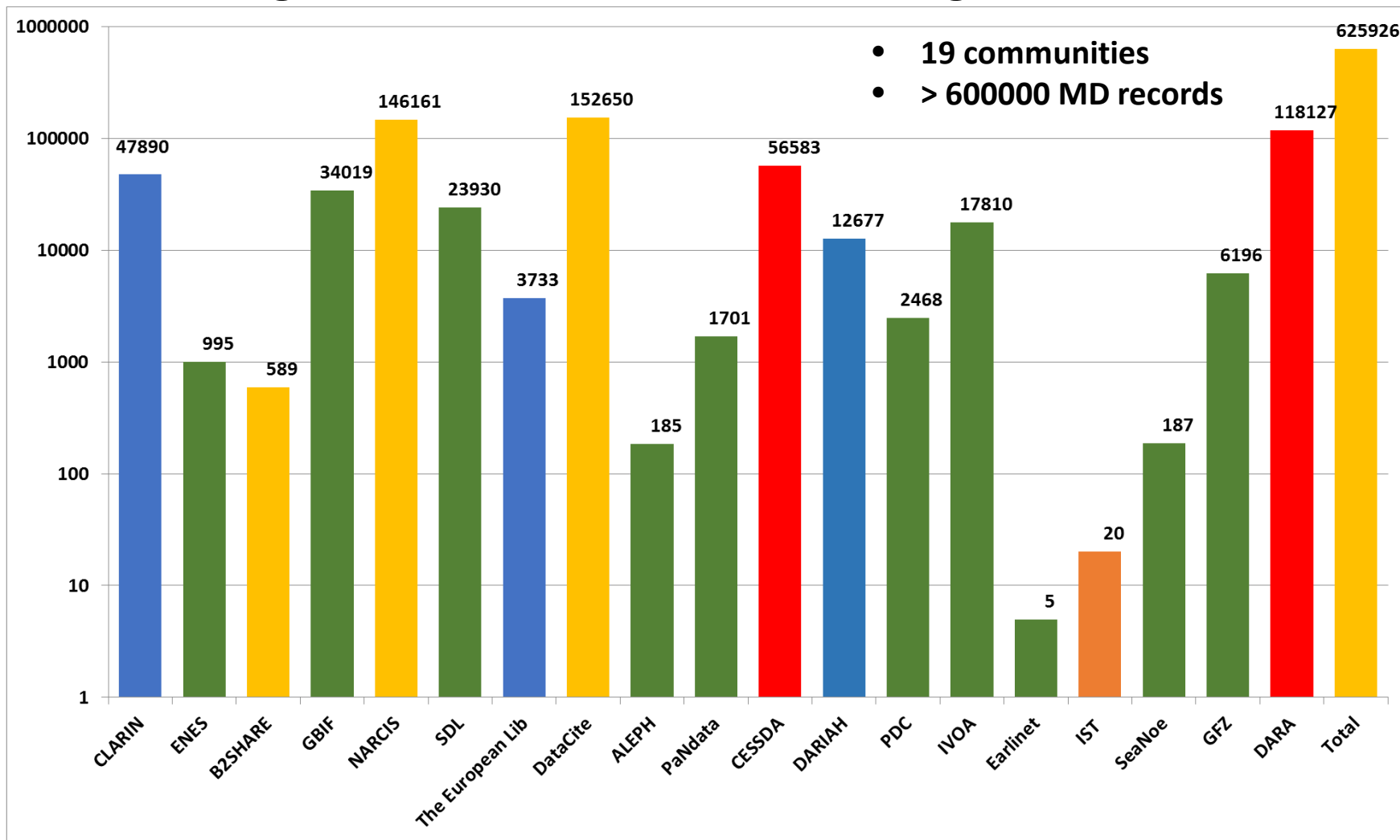


Metadata Type	B2FIND Field name	Allowed values	Semantic definition	Level of Obligation	Occurrence
General information	Title	Free text (unicode)	A name or title a resource is known	Mandatory	1
	Description	Free text	Additional info	Recommended	0-1
Data Access	Source	Valid URL or URN	Unique link to data resource	Mandatory (1)	0-1
	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

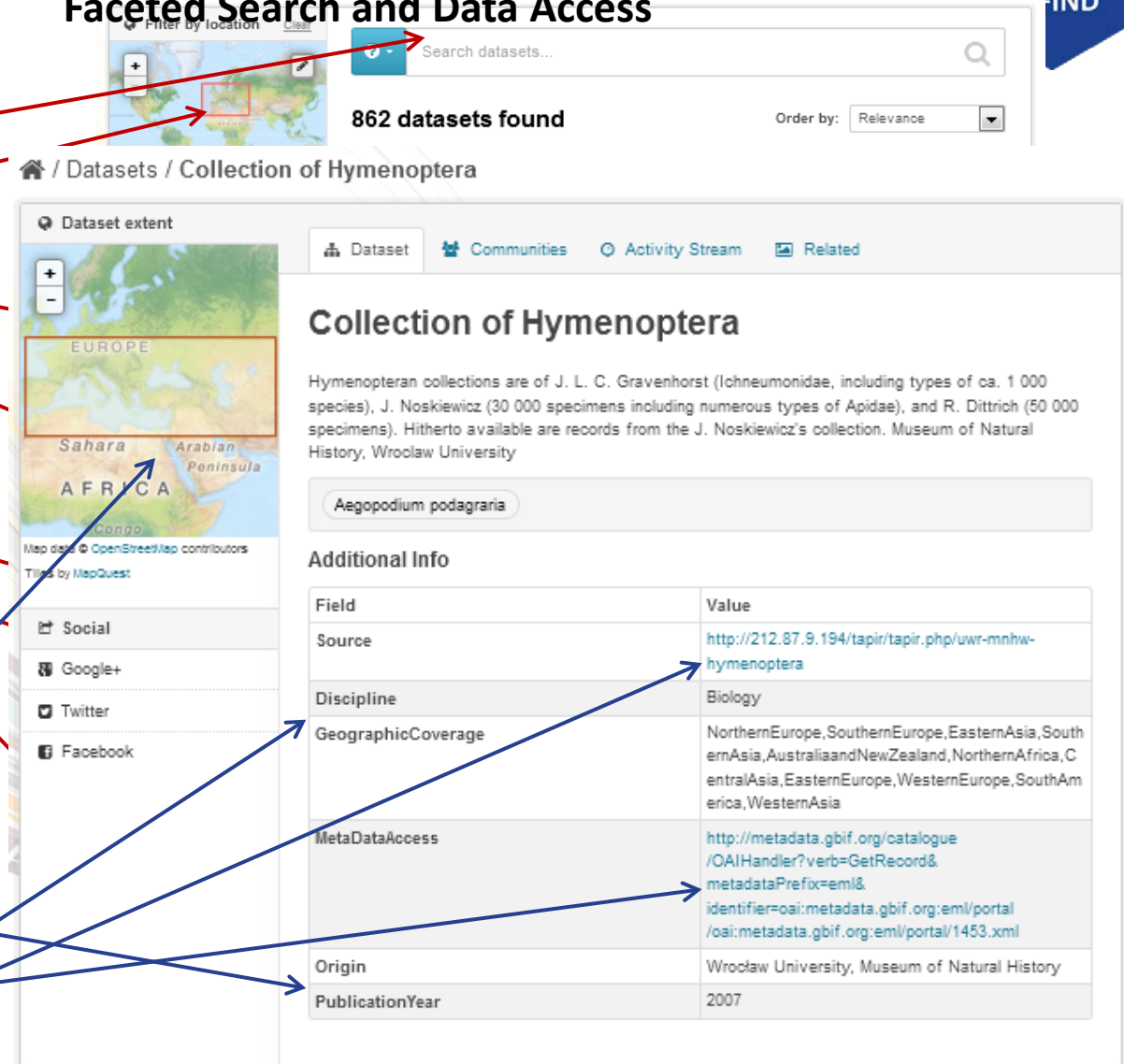
Ingestion Status of the MD Catalogue



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.



The screenshot shows the B2FIND Discovery Portal interface. At the top, there is a search bar with the text "Search datasets..." and a magnifying glass icon. Below the search bar, it says "862 datasets found" and "Order by: Relevance". The main content area is titled "Collection of Hymenoptera" and includes a map of Europe and Africa with a red box highlighting the region. To the right of the map is a table of field-value pairs.

Field	Value
Source	http://212.87.9.194/tapir/tapir.php/urw-mnhw-hymenoptera
Discipline	Biology
GeographicCoverage	NorthernEurope, SouthernEurope, EasternAsia, SouthernAsia, AustraliaandNewZealand, NorthernAfrica, CentralAsia, EasternEurope, WesternEurope, SouthAmerica, WesternAsia
MetadataAccess	http://metadata.gbif.org/catalogue/OAIHandler?verb=GetRecord&metadataPrefix=eml&identifier=oai:metadata.gbif.org:eml/portal/oai:metadata.gbif.org:eml/portal/1453.xml
Origin	Wroclaw University, Museum of Natural History
PublicationYear	2007

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>

Data provider
(Community)



Harvest specification :

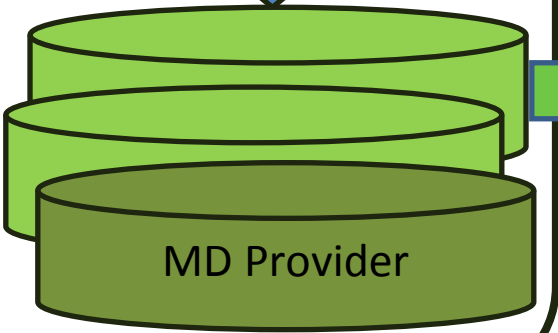
- OAI-URL
- OAI subsets
- MD formats

Mapping specification :

- XPATH rules
- Community specific MD schemas and ...

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

MD Generation and Specification



MD Harvesting

Mapping and Validation

Uploading and Indexer

User
(Scientist or Researcher)



Search and Data Access

EUDAT-B2FIND

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