

# **EUDAT B2FIND**

## **A Cross-Discipline Metadata Service and Discovery Portal**

Heinrich Widmann, DKRZ  
DI4R 2016,  
Krakow, 28 September 2016



[www.eudat.eu](http://www.eudat.eu)

# Outline

- EUDAT and the B2 Service Suite
- Guidelines and Concepts
- B2FIND – EUDAT’s Discovery Service
  - MD Ingestion and the B2FIND Schema
  - Disciplines, Communities and the MD catalogue
  - Data Access Identifiers
  - Discovery Portal
- Outlook and Summary

# EUDAT and the B2 Service Suite

- The project European Data Infrastructure (EUDAT)
  - funded by the EU Horizon2020 program
  - started in 2011, now in 2<sup>nd</sup> phase 'EUDAT2020', will end 2018
  - $\geq 2018$  : agreement of cooperation
- Motivation : Manage the rising tide of research data
  - Improve Interoperability in a wide cross-disciplinary scope
- Objective : Build up a Collaborate Data Infrastructure,
  - based on common data services
  - driven by requirements of the research communities

# B2 Service Suite

→ <http://www.eudat.eu/services>



## B2DROP

Sync and Exchange Research Data



## B2SHARE

Store and Share Research Data



## B2SAFE

Replicate Research Data Safely



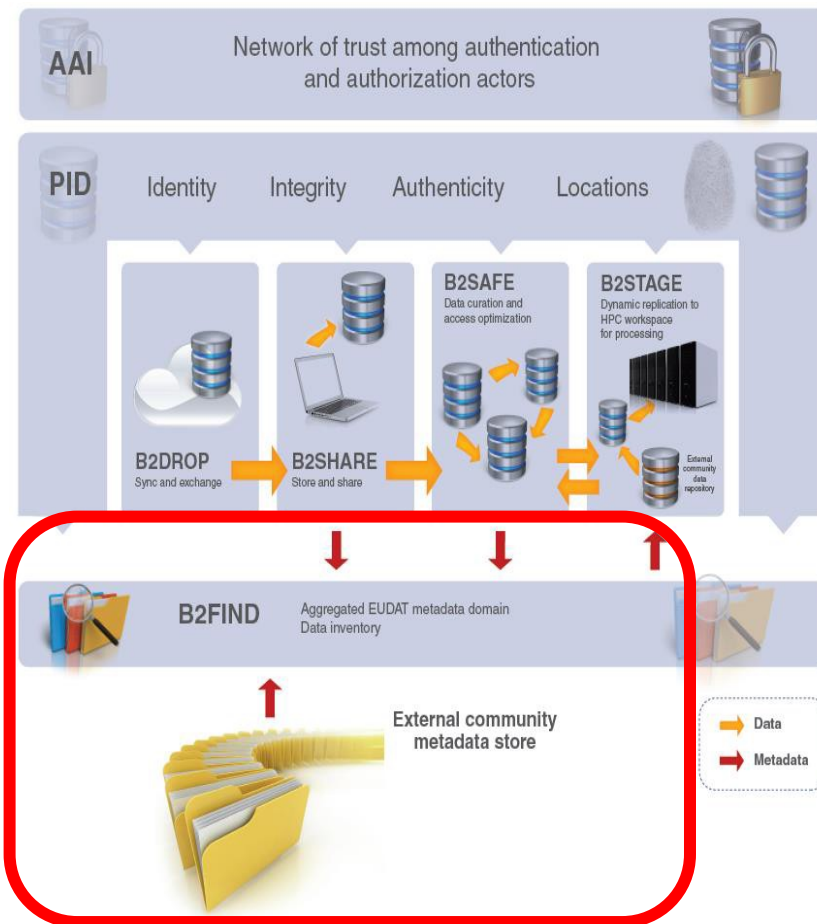
## B2STAGE

Get Data to Computation



## B2FIND

Find Research Data



# Guidelines and Concepts

- **The FAIR principles**
  - **B2FIND approach**

- **Findability** := “Ease with which information can be found”
  - Powerful and easy-to-use search features and functionalities
- **Accessibility** := “Ability to access [ ] data stored within repositories”
  - Unique and persistent identification and resolvability of data objects
- **Interoperability** : “Ability of multiple systems with different [ ] structures to exchange data with minimal loss of content [ ]” (NISO)
  - Comprehensive cross-disciplinary MD catalogue based on common standards and by minimising loss of information
- **Reuseability** := “Ability to re-use data created by others”
  - Cross-discipline approach and catalogue covering multiple sources

# Levels of Interoperability

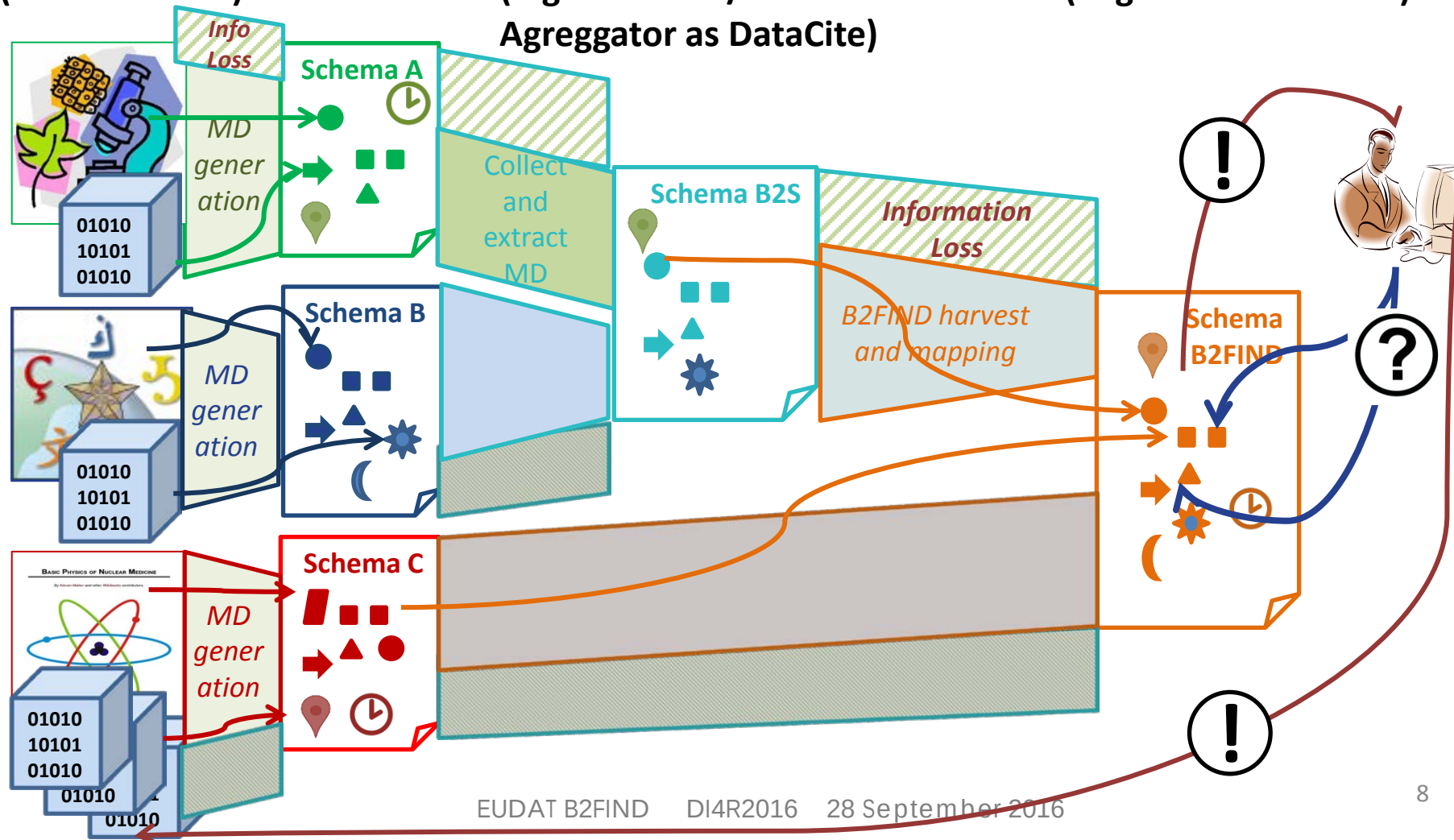
Heterogeneity

Homogeneity

Research Communities  
(Data Provider)

Data Repositories  
(e.g. B2SHARE/B2SAVE or  
Agreggator as DataCite)

Service Provider  
( e.g. EUDAT-B2FIND )





# B2FIND

## MD Ingestion and Common Schema

# B2FIND Ingestion Workflow

**Data provider  
(Community)**



Harvest specification :

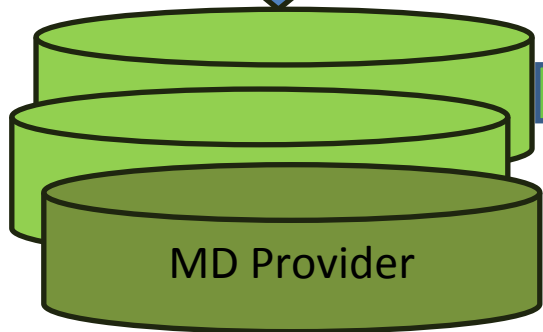
- OAI-URL
- OAI subsets
- MD formats

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
- Guarantee 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

# B2FIND MD Schema (extract)

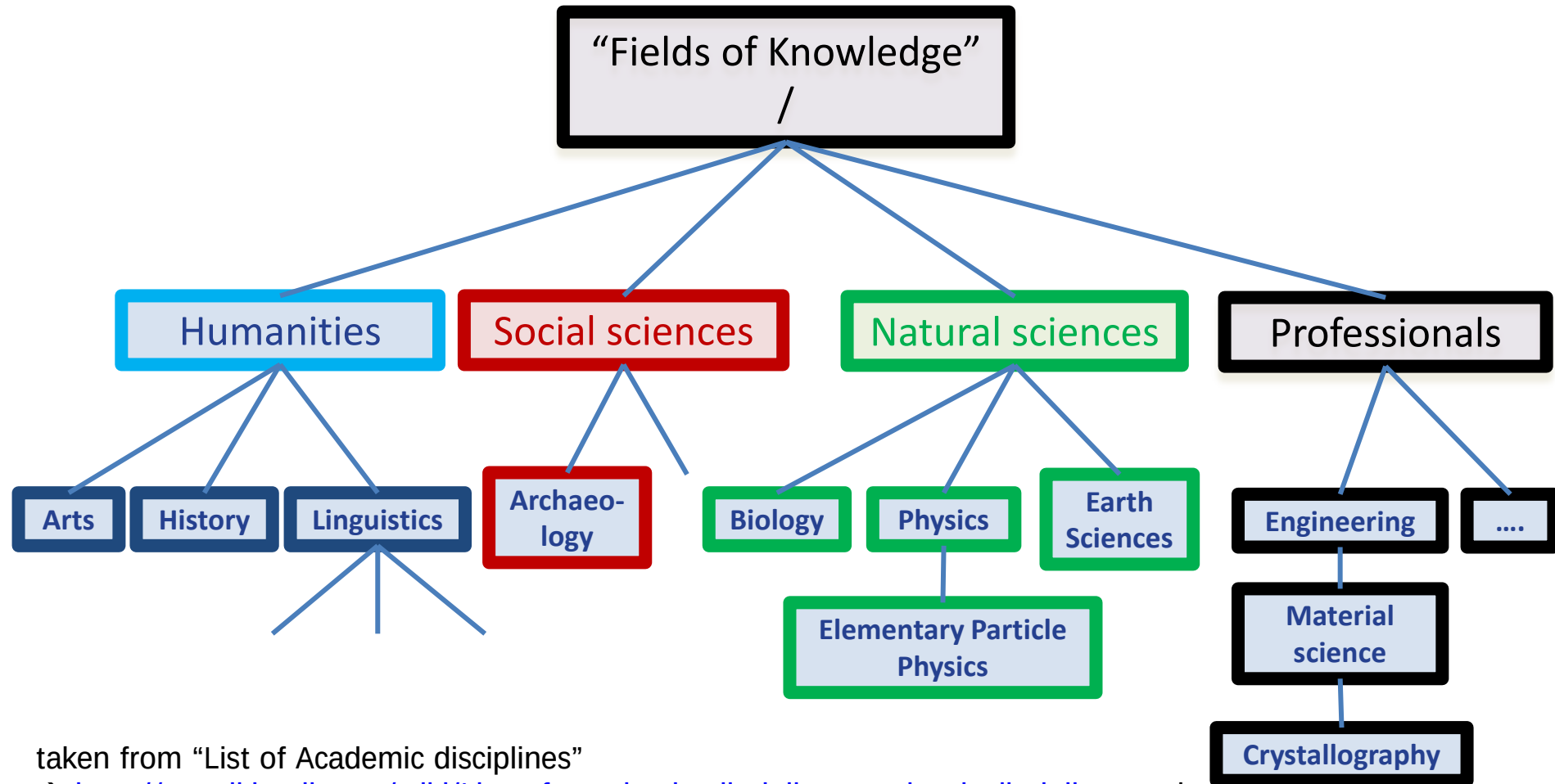
| Metadata Type       | B2FIND Field name | Allowed values                            | Semantic definition                     | Level of Obligation | Occurrence |     |
|---------------------|-------------------|-------------------------------------------|-----------------------------------------|---------------------|------------|-----|
| 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        | 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        |     |

# B2FIND

## Disciplines, Communities and MD Catalogue

# The Facet „Discipline“

## Controlled Vocabulary



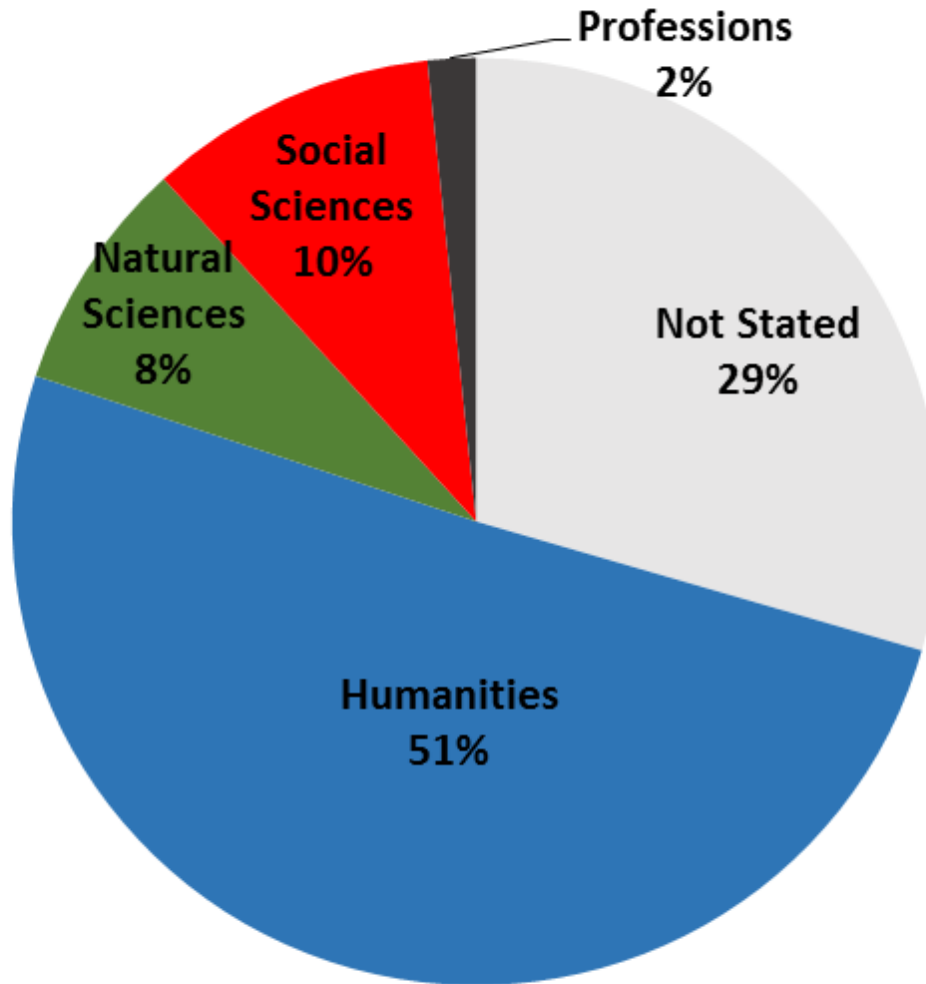
taken from "List of Academic disciplines"

→ [http://en.wikipedia.org/wiki/List\\_of\\_academic\\_disciplines\\_and\\_sub-disciplines](http://en.wikipedia.org/wiki/List_of_academic_disciplines_and_sub-disciplines) and

„The Fields of Knowledge“

→ [http://www.thingsmadethinkable.com/item/fields\\_of\\_knowledge.php?focus=natural\\_sciences](http://www.thingsmadethinkable.com/item/fields_of_knowledge.php?focus=natural_sciences)

# Coverage of Disciplines in B2FIND



# B2FIND MD Catalogue

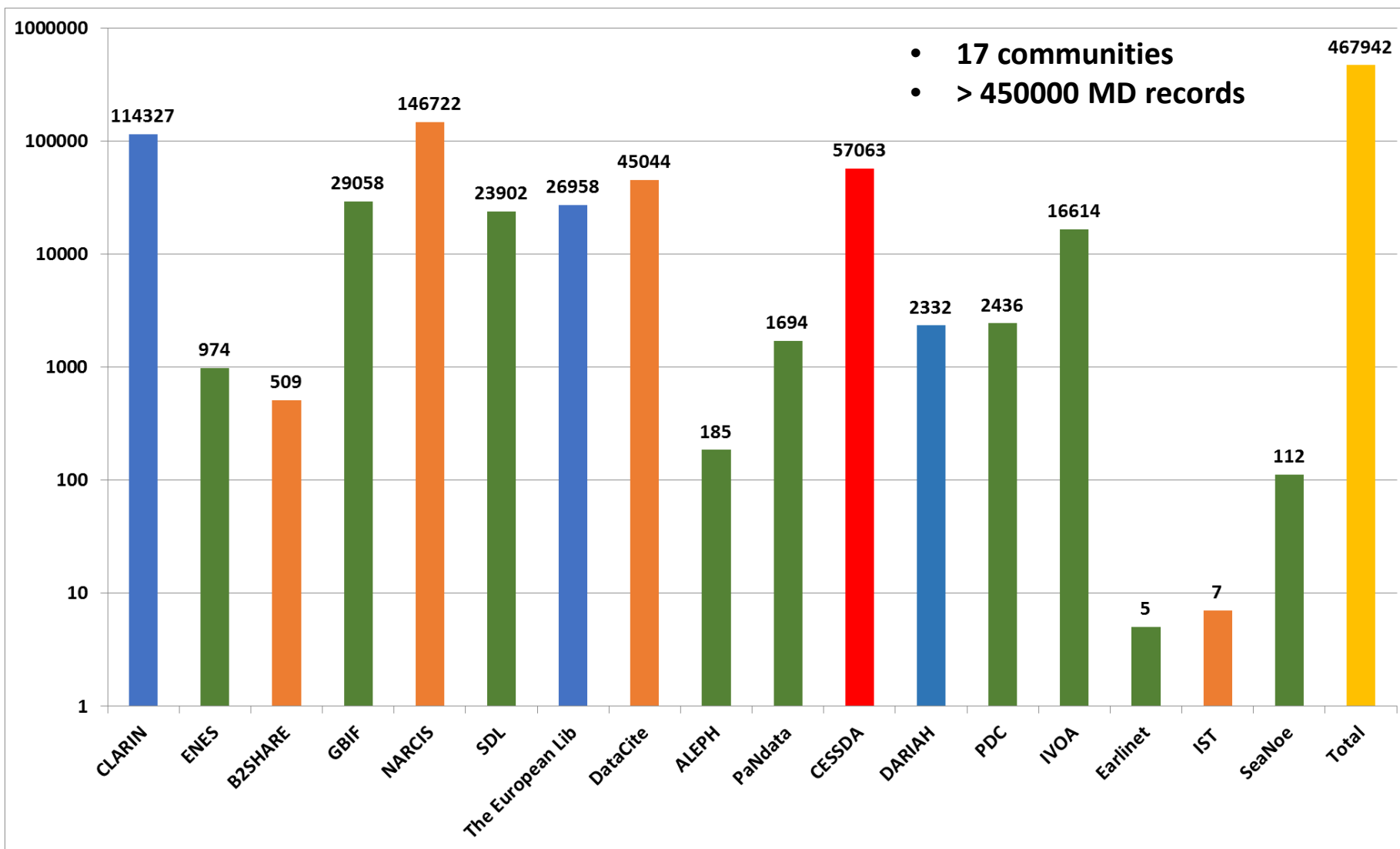
Ingestion status

Humanities

Social Sciences

Natural Sciences

Cross Discipline



# B2FIND

## Data Access



# Data Access Identifiers

## Resolvability and 'Levels of aggregation'

B2FIND

Resource

XML

```
</dc:identifier
value>
```

</>



B2FIND  
Metadata

Source

PID

DOI

Resolution and Access

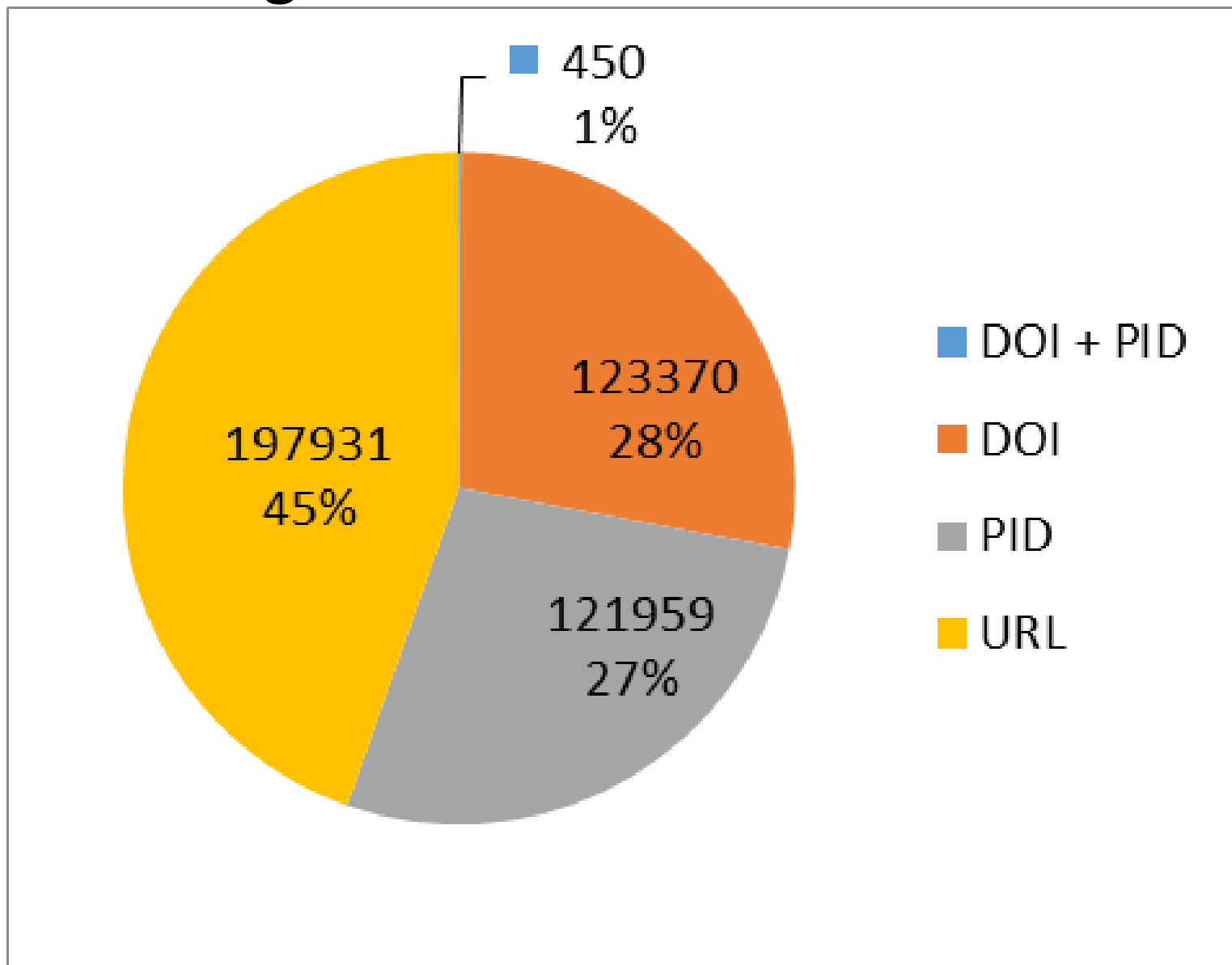
| Type            | Unique | Persistent | Resolvable | Citable |
|-----------------|--------|------------|------------|---------|
| DOI             | ✓      | ✓          | ✓          | ✓       |
| PID             | ✓      | ✓          | ✓          | X       |
| URL<br>(Source) | ✓      | ?          | ?          | X       |

Data Collection

01010  
10101  
01010

Stricter  
Policies

# Coverage of Data Access Identifiers



# B2FIND

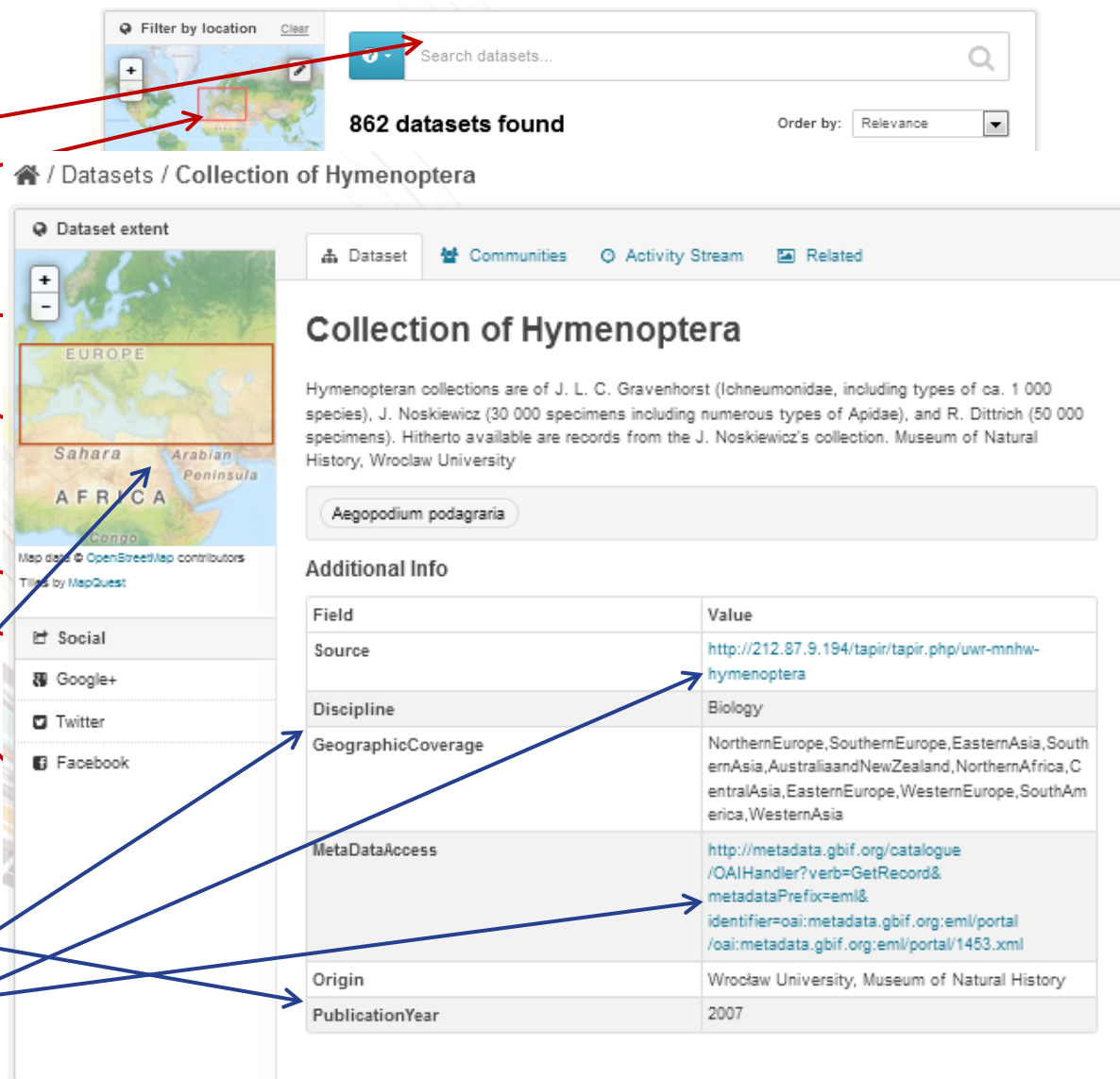
## Discovery Portal

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



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'. To the left of the search bar is a map with a red box highlighting a region in Europe. Below the map, there is a 'Filter by location' section with a 'Clear' button. The main content area is titled 'Collection of Hymenoptera' and includes a description of the collection. Below the description, there is a table of field-value pairs. The table has two columns: 'Field' and 'Value'. The rows are: Source, Discipline, GeographicCoverage, MetadataAccess, Origin, and PublicationYear. The 'Value' column contains links and text for each field. The 'GeographicCoverage' field has a link to a map. The 'MetadataAccess' field has a link to a metadata page. The 'Origin' field has the text 'Wroclaw University, Museum of Natural History'. The 'PublicationYear' field has the text '2007'. The 'Social' section on the left includes links to Google+, Twitter, and Facebook. The 'Dataset extent' section on the left includes a map and a 'Map data © OpenStreetMap contributors' note.

Collection of Hymenoptera

Hymenopteran collections are of J. L. C. Gravenhorst (Ichneumonidae, including types of ca. 1 000 species), J. Noskiewicz (30 000 specimens including numerous types of Apidae), and R. Ditttrich (50 000 specimens). Hitherto available are records from the J. Noskiewicz's collection. Museum of Natural History, Wroclaw University

Aegopodium podagraria

Additional Info

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

# Outlook and Summary

# Outlook

- Handle scalability and granularity issues
  - 'Levels of aggregation'
- Metrics for Key Indicators and Metadata Quality
  - Establish content-related quality assurance
- Add further search and distribution channels, e.g.
  - Use linked data : Potential for semantic enrichment
  - 'Annotation' functionality : Users link datasets to external reference materials (vocabularies, ontologies, etc.)
  - Query-based Taxonomies : Enabling hierarchical search, e.g. in trees of 'Disciplines'

# Summary

- EUDAT-B2FIND
  - established an operative service based on agreed standards and guidelines as the FAIR principles,
  - provides a discovery portal with powerful search functionalities and
  - is based on a unique catalogue of research data , combining many heterogeneous and cross-discipline sources
- Improved interoperability is achieved by homogenisation to a common metadata schema
- Further efforts are made to address the demands of the communities and data projects, to adapt the system for future challenges

# Thank you for your attention !

Links :

- info : <http://eudat.eu/b2find>
- portal : <http://b2find.eudat.eu>

Contact

- [www.eudat.eu/support-request](http://www.eudat.eu/support-request)
- [widmann@dkrz.de](mailto:widmann@dkrz.de)

