

B2FIND enhancement with FAIRCORE4EOSC components

Heinrich Widmann (DKRZ)

Slides by Heinrich Widmann, Anna-Lena Flügel and Aenne Löhden



Funded by
the European Union

14. 12. 2023 | Heinrich Widmann | Enhancing Research Data Management with EUDAT and FAIRCORE4EOSC

Harvesting

- **OAI – PMH**
is preferred due to its possibility to define endpoints, mdprefixes, subsets
- **JSON – API**
direct JSON APIs are supported as well
- **CSW**
for harvesting from geonetworks
- **[SparQL]**
to collect RDFs and JSON-LD



Search

- **user-friendly discovery portal**
- **allows faceted search to narrow down results**
e.g. spatial/temporal coverage, facet „discipline“
- **reference-improvements planned**
e.g. ORCID, PIDs for instruments, PID graphs


Mapping

from community Schemes onto EUDAT Core Schema


- **Technical workflow**
Format converting, mapping onto B2FIND schema, semantic mapping using vocabularies
- **Intensive communication**
with communities to agree on content related issues and technical procedures
- **Supported Metadata Standards**
DublinCore, Datacite, DDI, CMDI, ISO 19115/139 [INSPIRE], MarcXML, but also community specific

Exchange with RDGraph & PIDGraph


- Integration of B2FIND as datasource in RD Graph
- Integration of PID Graph information in B2FIND’s discovery portal





RDGraph
EOSC Research
Discovery Graph


 EOSC | FAIRCORE4EOSC
Enabling a FAIR EOSC ecosystem


The **EOSC Research Discovery Graph** will:

 Provide regular data dumps and APIs to support data re-use


 Facilitate the discovery of products and services

 Enable community users to get recommendations for research product


 Enable discovery, monitoring and access to research products and services




PIDGraph
EOSC PID
Graph

 EOSC | FAIRCORE4EOSC
Enabling a FAIR EOSC ecosystem


The **EOSC PIDGraph** will:

 support service workflows for regularly generating data dumps related to community profiles for reuse by the EOSC community and related services


- Utilisation of MSCR (and DTR) for B2FIND by
- Uploading and updating source and target metadata schemas and crosswalks in the MSCR
 - Retrieving versions of schemas and crosswalks maintained in the MSCR





MSCR
EOSC Metadata Schema
& Crosswalk Registry


 EOSC | FAIRCORE4EOSC
Enabling a FAIR EOSC ecosystem

Expected impact of the **Metadata Schema and Crosswalk Registry (MSCR)**:


 Better metadata interoperability

 Easy to use GUI for creating crosswalks


 Incentivised FAIRification





DTR
EOSC Data
Type Registry

 EOSC | FAIRCORE4EOSC
Enabling a FAIR EOSC ecosystem

Expected impact of the **EOSC Data Type Registry (DTR) Component**:

 Registration of PID metadata elements as data types

 Facilitated search across existing data types

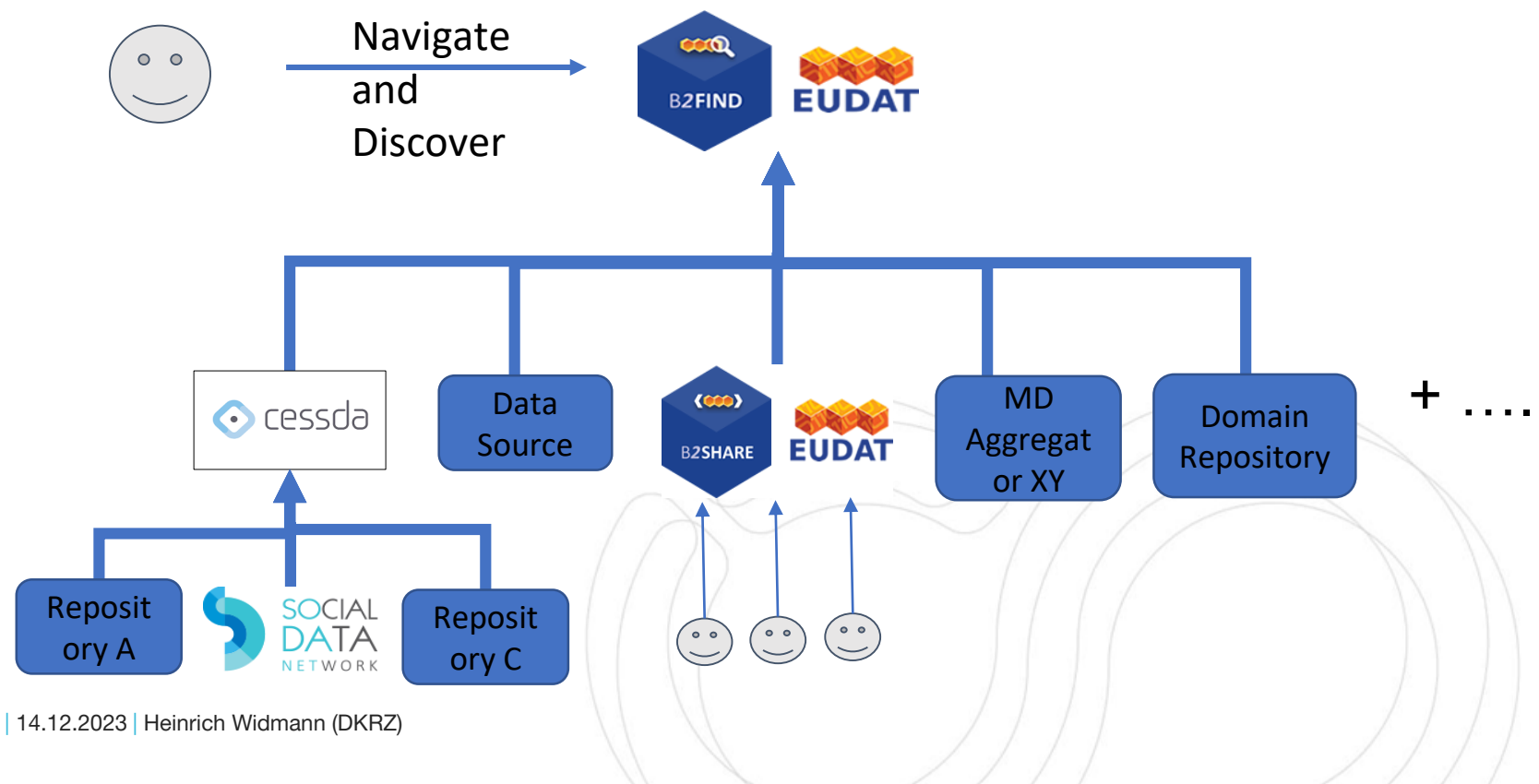
 Standardisation of PID metadata

- Enhanced cross-domain discovery of research products
 - on EOSC level by searching in the graphs and
 - via B2FIND discovery portal
 - Provision of provenance metadata in the RDGraph
 - Interlinkage to related information via the PIDGraph
-
- Aiming in good practises and guidelines for data sources and discovery services intend to utilise FC4E component

B2FIND

- harvests from ca. 100 data sources incl. repositories, metadata aggregators ...
- ... which maybe again harvest from (domain specific) repositories
- In particular metadata from EUDAT-B2SHARE are harvested continuously
- Inter-disciplinary approach and covering any kind of research discipline
- Sustainable, operable, hierarchical and incremental ingestion workflow

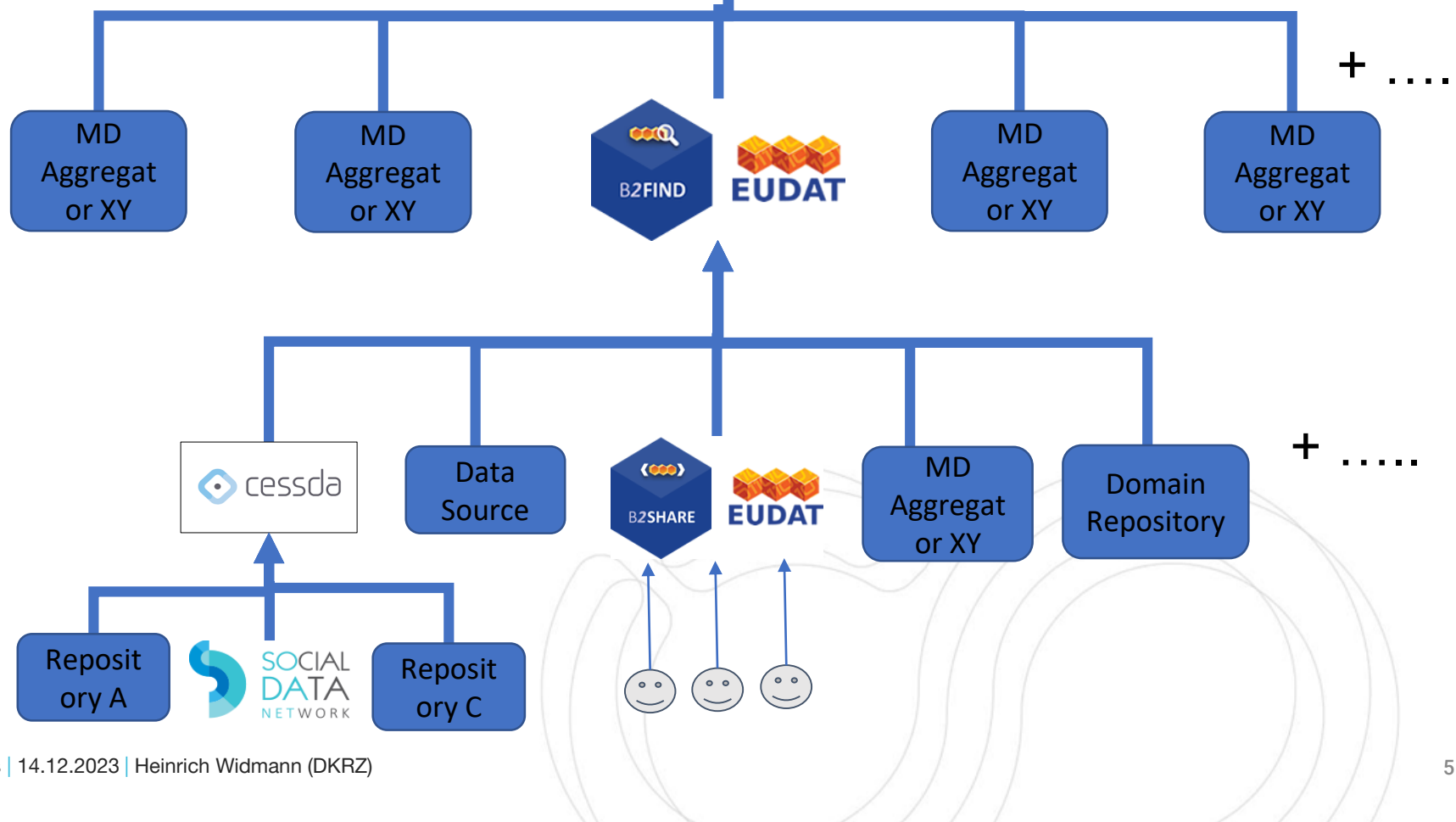
- allows interdisciplinary search for data resources covering a wide cross-domain field via B2FIND's discovery portal





Integration of B2FIND in EOSC RD Graph

- OpenAIRE harvests from B2FIND and hundreds of further data sources incl. Repositories, metadata aggregators ...
 - B2FIND follows OpenAIRE's guidelines for data providers
 - Keep and provide provenance information
 - RDGraph extended by B2FIND's comprehensive metadata
- yields in enhanced cross-domain discovery of 'research products' + provenance metadata via the EOSC RDGraph



Dataset are grouped under the data source
B2FIND has harvested from

Link to original harvested metadata record

Provenance metadata including

- Creator and Contact
- Publication information
- Rights, Open access tag
- And (if provided) instrument, ...

Organization

CESSDA
CESSDA provides large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe, with the aim of...

[read more](#)

Social

Twitter

Facebook

Dataset

Socio-Economic Development and Importance of Small-scale Manufacturing Enterprises in Malaysia

The present study examines the socio-economic development and importance of small manufacturing companies in Malaysia. For this purpose, Malaysian entrepreneurs were interviewed with regard to their professional biography, their working conditions, the type of their business, as well as the business-related day-to-day work and the financial situation of their company.

Die vorliegende Studie untersucht die sozioökonomische Entwicklung und Bedeutung von kleinen Produktionsunternehmen in Malaysia. Zu diesem Zwecke wurden malaysische Kleinunternehmer hinsichtlich ihrer Berufsbiografie, ihrer Arbeitsbedingungen, der Art ihres Unternehmens sowie des damit zusammenhängenden Berufsalltags und der finanziellen Situation ihres Unternehmens befragt.

Economic systems an... **HANDEL** **INDUSTRIE UND MÄRKTE** **INDUSTRY AND MARKETS**

TRADE Wirtschaftssysteme ...

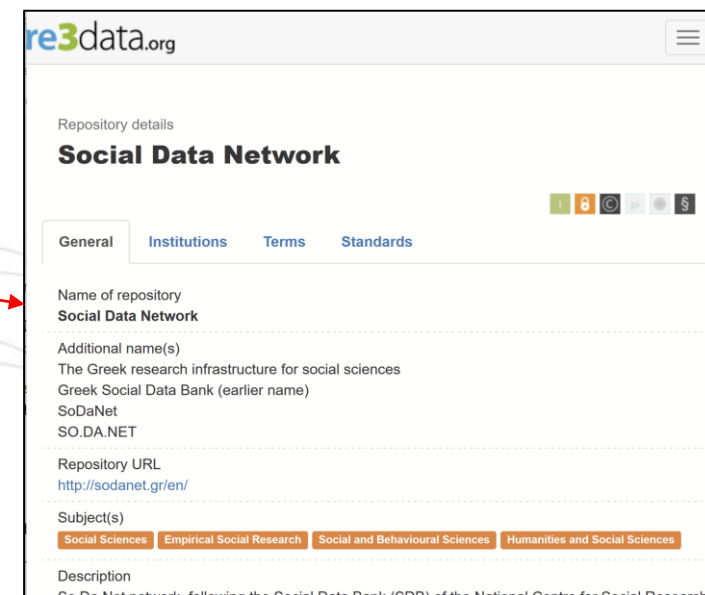
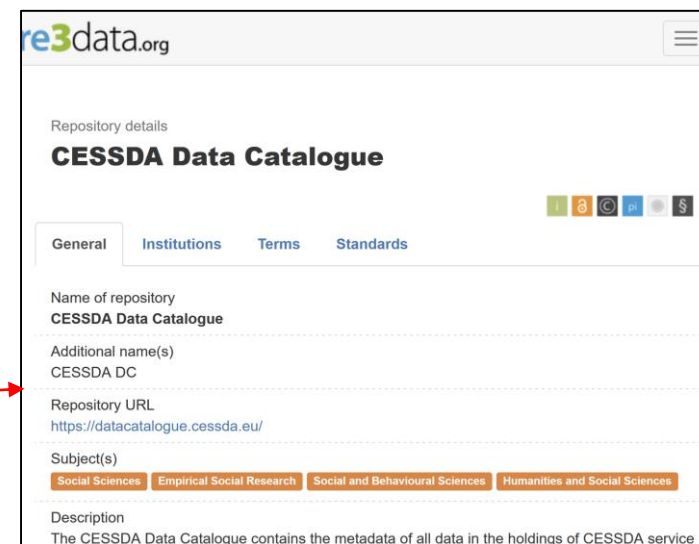
Identifier	
Source	https://search.gesis.org/research_data/ZA5039?lang=de
Metadata Access	https://datacatalogue.cessda.eu/oai-pmh/v0/oai?verb=GetRecord&metadataPrefix=oai_ddi25&identifier=fd300e63ded90859bb347c9bfde15f4e15a490ebd61887f0e86e50586243ad9d

Provenance	
Creator	Borsutzky, Dieter
Publisher	GESIS Data Archive for the Social Sciences; GESIS Datenarchiv für Sozialwissenschaften
Publication Year	1991
Rights	B - Data and documents are released for academic research and teaching, if the results won't be published. If any publications or any further work on the results is planned, permission must be obtained by the Data Archive.; B - Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben, wenn die Ergebnisse nicht veröffentlicht werden. Sollte eine Veröffentlichung oder eine weitergehende Verarbeitung der Ergebnisse geplant sein, ist eine Genehmigung über das Datenarchiv einzuholen.
OpenAccess	true
Contact	http://www.gesis.org/

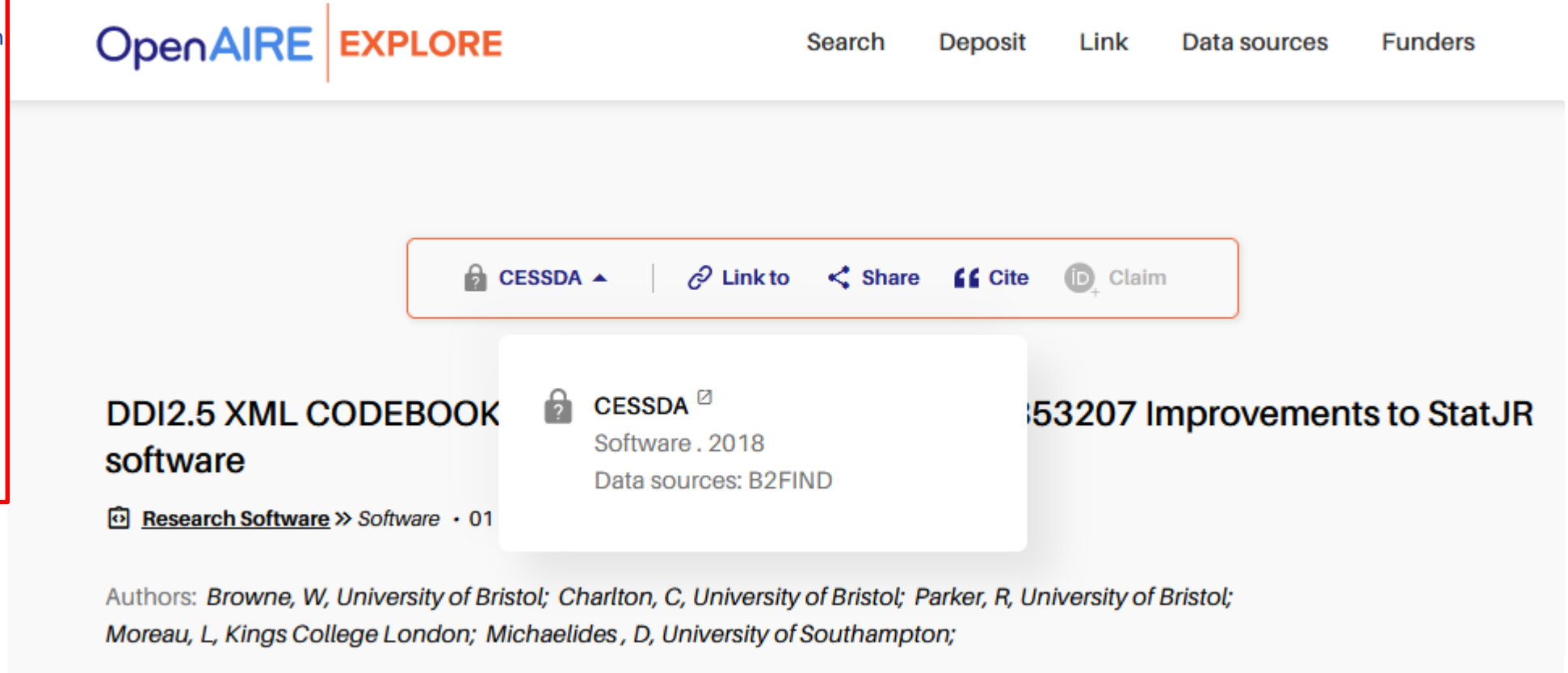
B2FIND harvests CESSDA, which itself harvests So.Da.Net



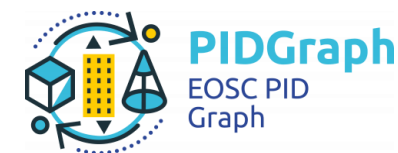
```
<about>
  <provenance ...>
    <originDescription harvestDate="later..." ...>
      <baseURL>http://...</baseURL>
      <repositoryID>re3data:r3d100010202</repositoryID>
      <repositoryName>CESSDA</repositoryName>
      <identifier>oai:...</identifier>
      <timestamp>later...</timestamp>
    ...
    <originDescription harvestDate="earlier..." ...>
      <baseURL>https://datacatalogue.sodanet.gr/oai</baseURL>
      <identifier>oai:...</identifier>
      <timestamp>earlier</timestamp>
    ...
      <repositoryID>re3data:r3d100010134</repoID>
      <repositoryName>So.Da.Net</repoName>
    </originDescription>
  </provenance>
</about>
```



OpenAIRE harvests B2FIND, which itself harvests CESSDA

- Use PIDGraph to interlink resource to related data, tools, instruments, projects, persons, ...
 - RDGraph is extended by PIDGraph
 - Datasets with DOI get integrated in the PIDGraph
 - B2FIND assigns standardized identifiers allowing referring to external information
- User get redirected and referenced to external information and entities

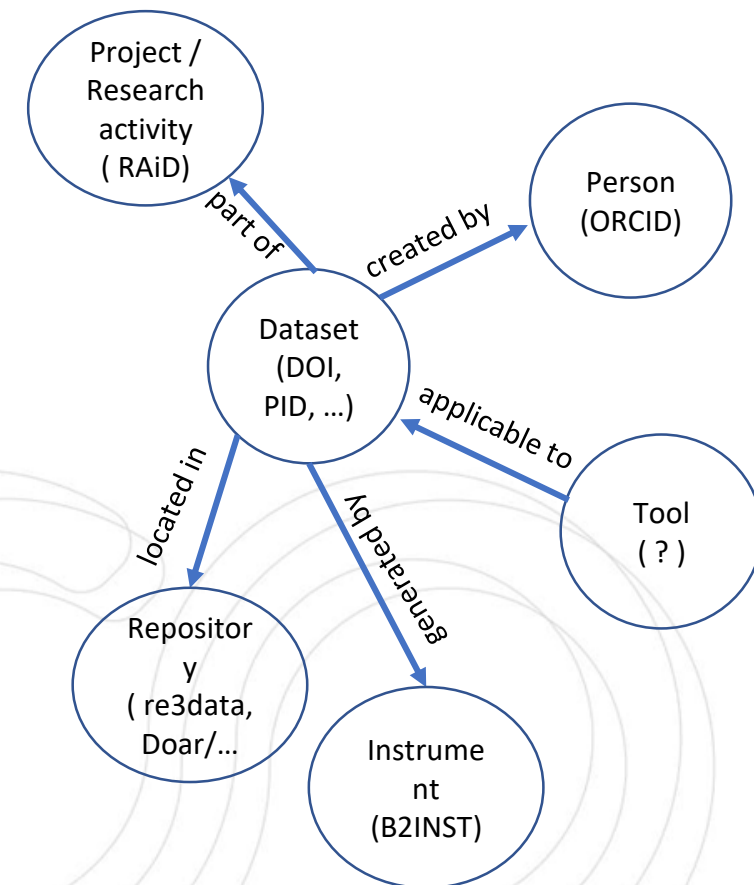
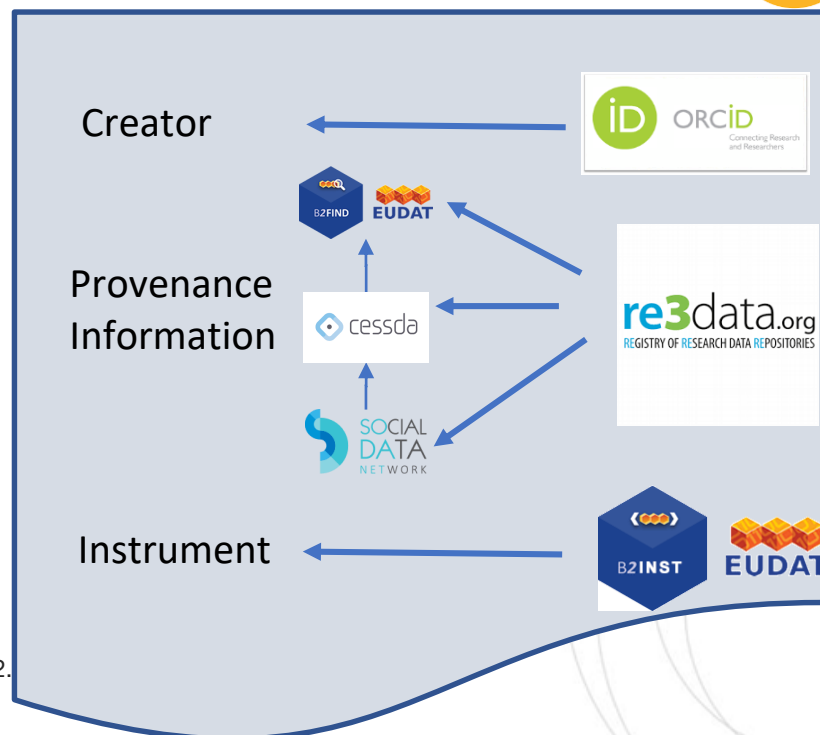


Resource, Objects,
Persons, ...

identifies

Standardised and
persistent Identifier

B2FIND
dataset



B2FIND dataset with Instrument

[DATA CATALOGUE](#)
[REPOSITORIES](#)
[PROJECTS](#)
[ABOUT](#)

[Dataset / HDAP -- Heidelberg...](#)

Organization

Dataset

Provenance

Creator	Mandel, H.; Birkle, K.; Demleitner, M.; Landessternwarte Heidelberg
Instrument	Bruce Double Astrograph, ; Calar Alto Schmidt Telescope, ; Calar Alto 1.23m Cassegrain, ; 72cm Walz Reflector; Calar Alto 2.2m Cassegrain; Calar Alto 3.5m 3-lens corrected; Calar Alto 3.5m 2-lens corrected; 2.2m MPG telescope at La Silla; Wolf's 15cm Doppelastrograph Königstuhl; Wolf's 15cm Doppelastrograph Märzgasse
Publisher	The GAVO DC team
Publication Year	2007
Rights	If you use this data, please acknowledge: "This work made use of the HDAP which was produced at Landessternwarte Heidelberg-Königstuhl under grant No. Heidelberg

Instrument	Bruce Double Astrograph, ; Calar Alto Schmidt Telescope, ; Calar Alto 1.23m Cassegrain, ; 72cm Walz Reflector; Calar Alto 2.2m Cassegrain; Calar Alto 3.5m 3-lens corrected; Calar Alto 3.5m 2-lens corrected; 2.2m MPG telescope at La Silla; Wolf's 15cm Doppelastrograph Königstuhl; Wolf's 15cm Doppelastrograph Märzgasse
Publisher	The GAVO DC team

[HELP](#)
[COMMUNITIES](#)
[REGISTER](#)
[CONTACT](#)

[Login](#)

[RECORDS](#)
C02620725258417A96C2428CA2D03341

Latest Version - Mar 21, 2023

Bruce Double Astrograph

PID: 21.T11975/6447eba9-caa4-4edf-a050-3ebf11b9609d

Basic metadata

Owners	Owner Name	Landessternwarte Heidelberg
Manufacturers	Manufacturer Name	:tba
Instrument Types	Instrument Type Name	Telescope
Open Access	True	✓
Instrument Identifier	Identifier	:unas
	identifierType	:unas
Landing Page	:unap	
Schema Version	1.0	

[HELP](#)
[COMMUNITIES](#)
[REGISTER](#)
[CONTACT](#)

[Login](#)

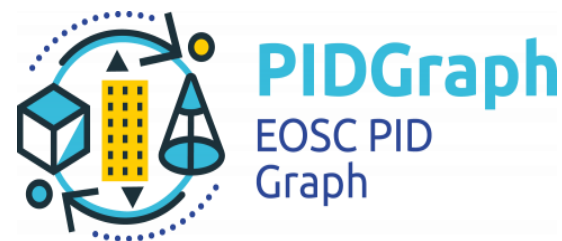
Register and publish your scientific instruments

Search for scientific instruments or register as a user to register and publish your own instrument!

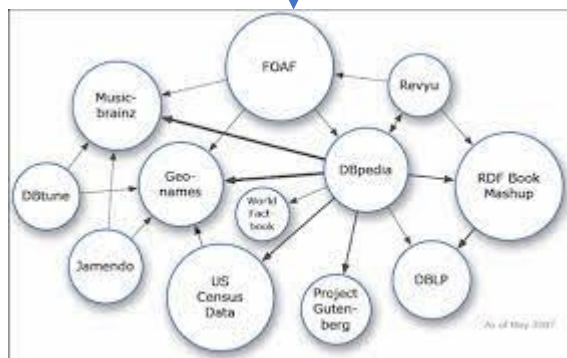
<https://b2inst.gwdg.de/>

- Retrieve graph dump via PID Graph REST API
- extend B2FIND records by links and references to information like
 - open access
 - related publications
 - instruments
 -
- yields in interlinking and redirecting to external related data, information or ontologies etc.

Work on this has just started with first ideas and using the recently launched PIDGraph API ...



Retrieve
/ harvest
graph
dumps



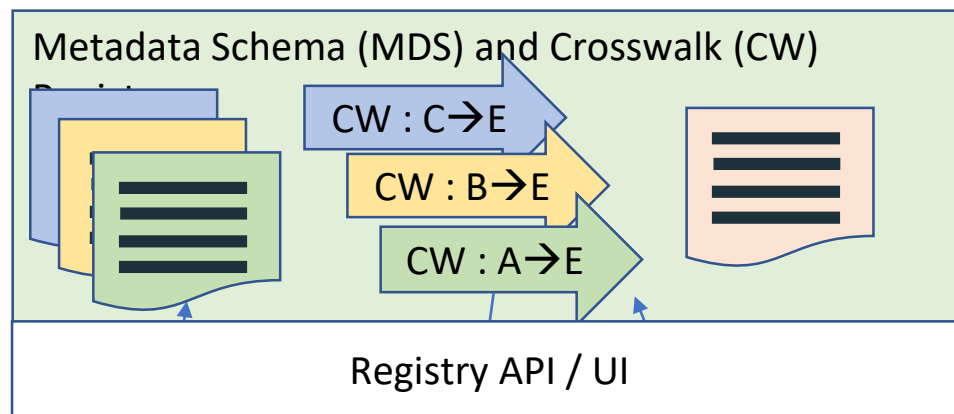
Enhance and extend B2FIND with
additional links/references/LoD



External
Graph A

B2FIND utilizes MSCR to support the operational ingestion workflow

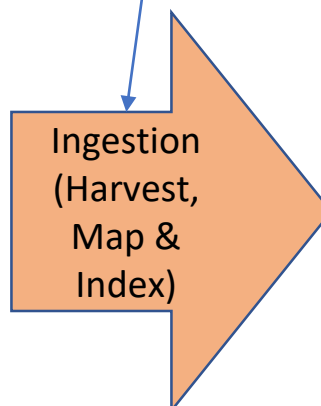
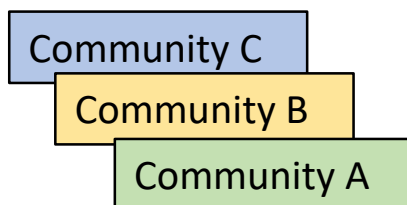
1. B2FIND registers
 - a. community (source) MDSs
 - b. EUDAT Core (target) MDS
2. Communities update ,their' MDSs (add specific fields, ...)
3. B2FIND registers crosswalk (CW) from source to target MDS in MSCR
4. MSCR maintains versions of MDSs and CWs
5. B2FIND retrieves up-to-date MDSs and CWs to use them for operational ingestion workflow



Update
Community
MDS

Retrieve
and
Utilise
CWs

Register source
and target MDS
and CWs



EUDAT Core
MDS E

MSCR
EOSC Metadata Schema & Crosswalk Registry
Enabling a FAIR EOSC ecosystem

Expected impact of the **Metadata Schema and Crosswalk Registry (MSCR)**:

- Better metadata interoperability
- Easy to use GUI for creating crosswalks
- Incentivised FAIRification

- Central registration, maintenance and versioning allows interoperable access to and exchange of up-to-date MD schemas and crosswalks, simplifying community intake and operational management.

EUDAT info and helpdesk

info@eudat.eu

helpdesk@eudat.eu

B2FIND Portal

<http://b2find.eudat.eu/>

B2FIND Guidelines for data provider

<http://b2find.eudat.eu/guidelines/introduction.html>

