



Interdisciplinary discovery in practice

B2FIND



eosc-hub.eu



[@EOSC_eu](https://twitter.com/EOSC_eu)



- OAI – PMH
- JSON – API
- CSW
- [SparQL]

Harvesting

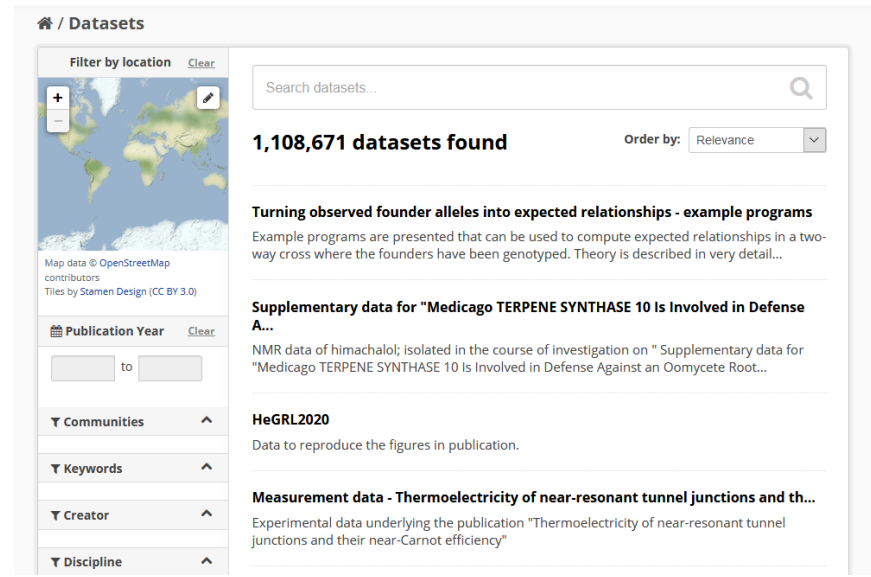


Search

- free text search
- faceted search
 - e.g. spatial/temporal coverage
- narrow down results

Mapping

- **technical** format conversion
- **semantic** reader for standards, mapping to target schema, including vocabularies, extensive validation
- **specific** due to Communities' needs

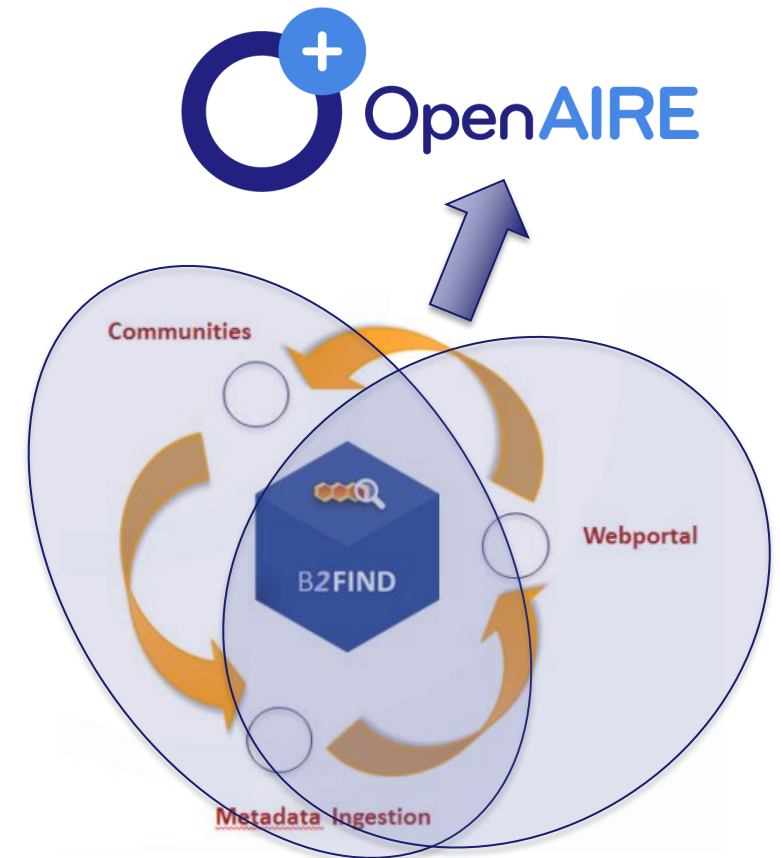


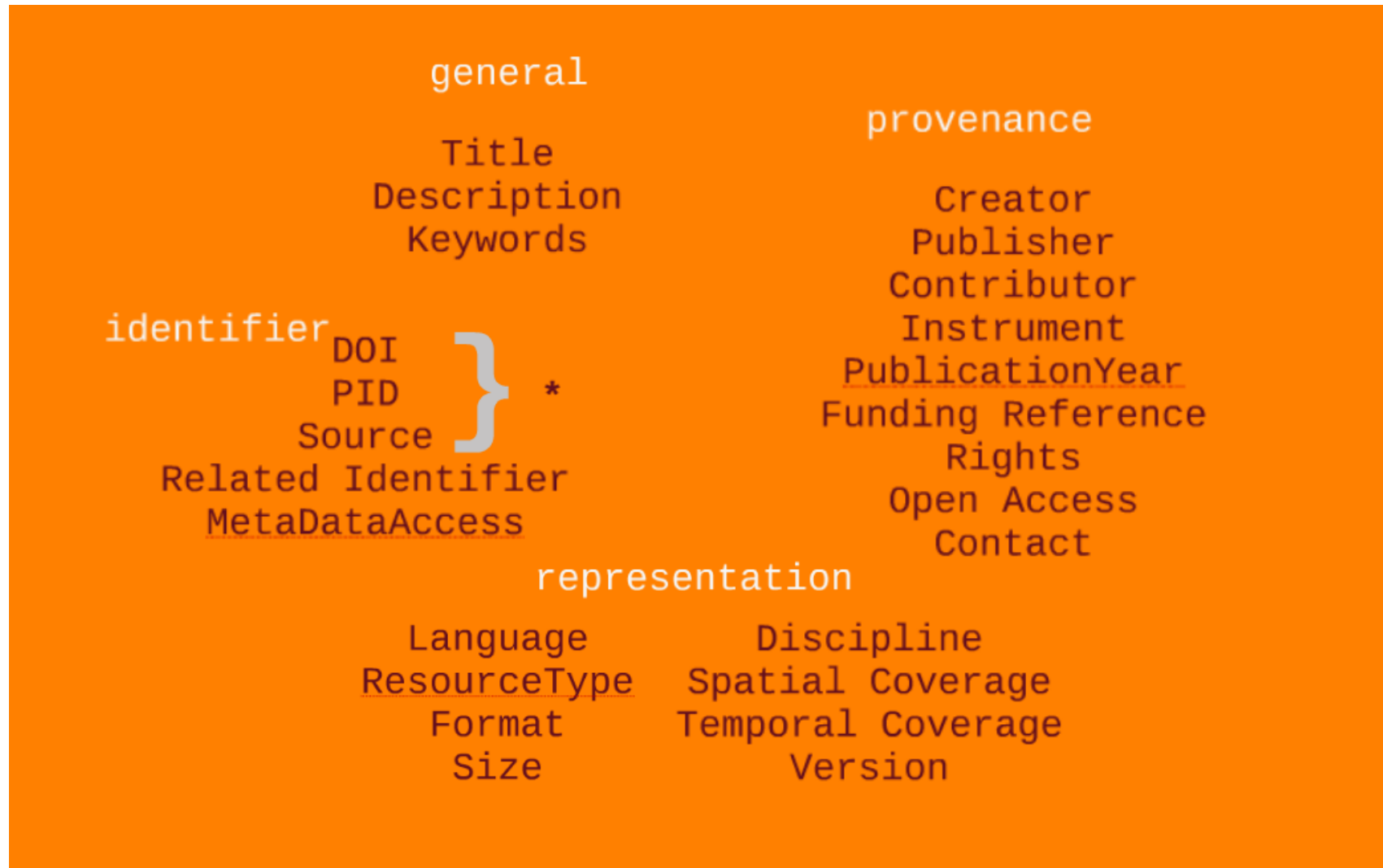
- Reader
Datacite, DublinCore, ISO19139, FGDC, FF (DDI and CMDI in old version)
- Community specific mapfiles -> curation!

- new **physical machine** (blades), new test VMs
- new **software stack**:
 - B2FIND ingestion code v3.0
<https://github.com/EUDAT-B2FIND/md-ingestion/releases/tag/v3.0.0>
 - CKAN interface
<https://github.com/EUDAT-B2FIND/ckanext-b2find>
 - Metadata exposure
<https://github.com/EUDAT-B2FIND/ckanext-oaipmh-server>
- new **B2FIND metadata schema v2.0**
 - XSD implementation: <http://b2find.eudat.eu/schema/b2f/2.0/meta.xsd>
 - Schema in Guidelines: <http://b2find.eudat.eu/guidelines/mapping.html>



- ◉ **B2FIND ingestion code v3.0**
 - very flexibel
 - allows harvesting from different endpoints within one Community
 - allows specific mapping for each endpoint within one Community
- ◉ **CKAN interface**
 - new facets
 - more upload options (bulk upload)
- ◉ **Metadata exposure**
 - CKAN extension for OAI-PMH
 - supported metadata prefixes: oai_b2f, oai_datacite





why yet another generic metadata schema?

➔ because we need it...

based on Datacite additional elements

- Community
- Meta Data Access
- Instrument
- Temporal Coverage

◎ B2FIND

- Community
- Identifier
either DOI /PID / Source
- Discipline
- Instrument
- Contact
- TemporalCoverage
- SpatialCoverage

◎ Datacite

- / (prefix)
- only DOI
<alternateIdentifier>
- Subject
- /
- within <contributorType
= ContactPerson>
- Date
- GeoLocation

◎ OpenAire

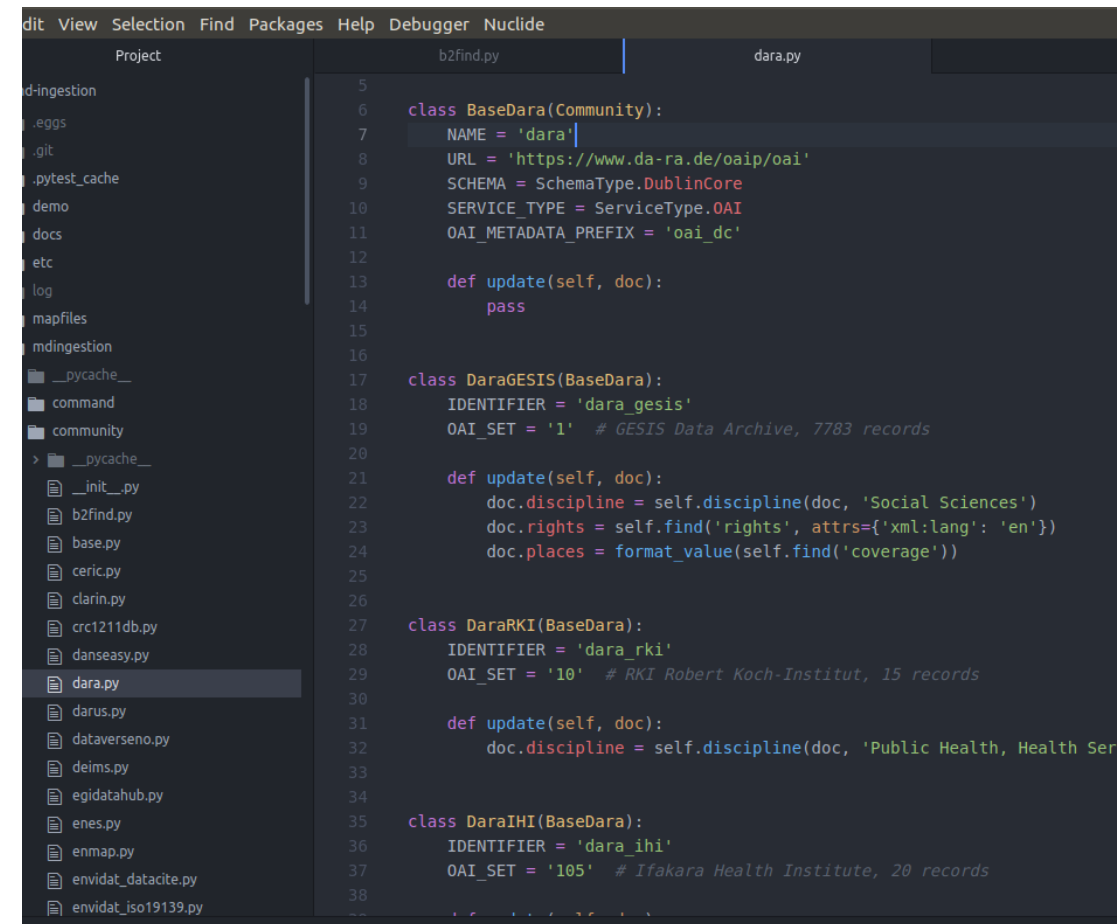
- Identifier with
different types (DOI,
Handle, PURL, URL, URN,
ARK)
- Subject
- /
- within <contributorType
= ContactPerson>
- Date
- GeoLocation

how to find a balance between community-specific metadata that serve their needs on the one side and a metadata schema that is sufficiently generic to represent interdisciplinary research data but at the same time is specific enough to enable a useful search with satisfying search results?

```

dara_gesis: 1%|9 | 100/7798
dara_rki: 100%|#####| 15/15
dara_ihi: 100%|#####| 20/20
dara_dipf: 100%|#####| 903/903
dara_bifie: 100%|#####| 138/138
dara_zfkd: 100%|#####| 10/10
dara_zbw: 100%|#####| 156/156
dara_srda: 100%|#####| 2685/2685
dara_ebdc: 0%|
dara_iwh: 100%|#####| 7/7
dara_ahri: 100%|#####| 88/88
dara_rdc: 100%|#####| 1859/1859
dara_dzhw: 100%|#####| 28/28
dara_xhub: 100%|#####| 49/49
dara_manifesto: 100%|#####| 25/25
dara_dji: 100%|#####| 3/3
dara_iab: 50%|#####5 | 50/101
dara_deutschebundesbank: 100%|#####| 80/80
dara_lifbi: 100%|#####| 206/206
dara_csese: 100%|#####| 16/16
dara_gesisdatorium: 100%|#####| 231/231
dara_rwi: 100%|#####| 227/227
dara_shareeric: 100%|#####| 80/80
dara_icpsr: 100%|#####9| 39666/39667
dara_zpid: 100%|#####| 64/64
dara_hsrc: 100%|#####| 223/223
dara_soep: 100%|#####| 60/60
dara_dza: 100%|#####| 80/80
dara_iqb: 100%|#####| 102/102
dara_rdcbo1: 100%|#####| 18/18
dara_rdcbo2: 100%|#####| 2/2
dara_csda: 100%|#####| 475/475
dara_indepth: 100%|#####| 152/152

```



```

Project: md-ingestion
File: dara.py

5
6 class BaseDara(Community):
7     NAME = 'dara'
8     URL = 'https://www.da-ra.de/oaip/oaip'
9     SCHEMA = SchemaType.DublinCore
10    SERVICE_TYPE = ServiceType.OAI
11    OAI_METADATA_PREFIX = 'oai_dc'
12
13    def update(self, doc):
14        pass
15
16
17 class DaraGESIS(BaseDara):
18     IDENTIFIER = 'dara_gesis'
19     OAI_SET = '1' # GESIS Data Archive, 7783 records
20
21    def update(self, doc):
22        doc.discipline = self.discipline(doc, 'Social Sciences')
23        doc.rights = self.find('rights', attrs={'xml:lang': 'en'})
24        doc.places = format_value(self.find('coverage'))
25
26
27 class DaraRKI(BaseDara):
28     IDENTIFIER = 'dara_rki'
29     OAI_SET = '10' # RKI Robert Koch-Institut, 15 records
30
31    def update(self, doc):
32        doc.discipline = self.discipline(doc, 'Public Health, Health Serv
33
34
35 class DaraIHI(BaseDara):
36     IDENTIFIER = 'dara_ihi'
37     OAI_SET = '105' # Ifakara Health Institute, 20 records
38

```

<https://github.com/EUDAT-B2FIND/md-ingestion/blob/master/mdingestion/community/dara.py>

<http://b2find.eudat.eu/dataset?groups=dara>

- analysis of the *(meta)data catalogues* concept in different domain-specific research data infrastructures and research data repositories
- pilot proposal of metadata catalogue integration to improve cross-disciplinary FAIR uptake
- including EOSC Clusters (EOSC-Life, PaNOSC, ESCAPE, ENVRIFair, SSHOC)
- idea:
 - ✓ testing (meta)data catalogue integration from the clusters with **DCATv2** and **DDI-CDI**
 - ✓ using B2FIND as a **service provider**
 - ✓ (hopefully) improve interdisciplinary discovery

for B2FIND

- **implementing DDI-CDI and DCATv2 reader**
- **test B2FIND schema compability**

❖ Future projects

- clara.science = **classification for research areas**
Cooperation project with re3data to develop/revise a vocabulary for scientific disciplines, to be reusable
- PIDs for Instruments
within DICE, in cooperation with Datacite

❖ Discussion Points

- Linked Data or not Linked Data? [triplestore vs. key-value pairs]
- Granularity! [RDA Task Force]
- when is a Persistent Identifier really persistent? [policies vs. reality]



links

B2FIND portal

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

B2FIND Guidelines for data provider

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

B2FIND in GitHub

<https://github.com/EUDAT-B2FIND/md-ingestion>

B2FIND metadata schema concordance

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

B2FIND classification for disciplines

https://github.com/EUDAT-B2FIND/md-ingestion/blob/master/mapfiles/b2find_disciplines.json

Questions, comments, remarks?

contact

Claudia Martens

martens@dkrz.de

Anna-Lena Flügel

fluegel@dkrz.de