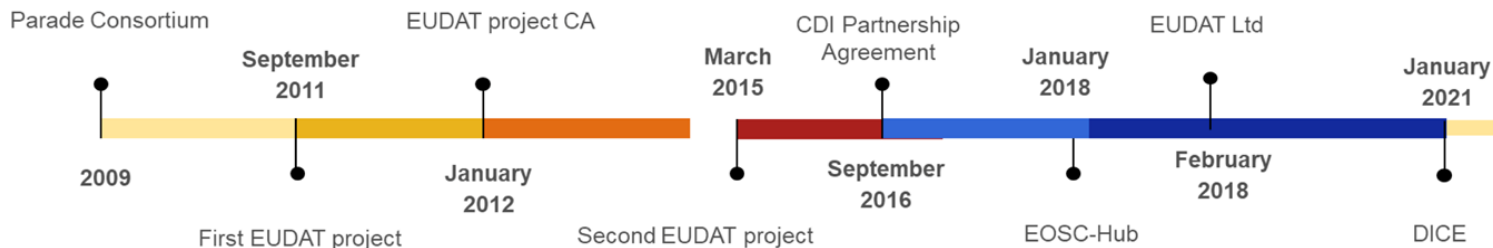


B2FIND

Discovery Portal for Research Data

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



- **already existing** simple and user-friendly **discovery portal** for **research data**
- joint metadata catalogue for records from very **divergent research areas** or **Infrastructures**
- central indexing tool for **EOSC-hub**, discovery service in **EOSC**

B2FIND – how does it work?

Harvesting

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

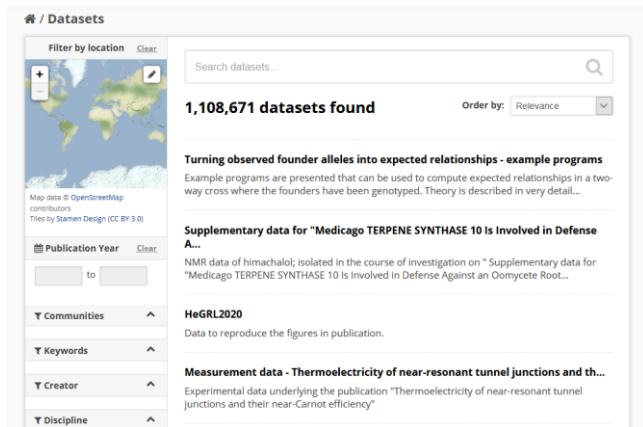


Search

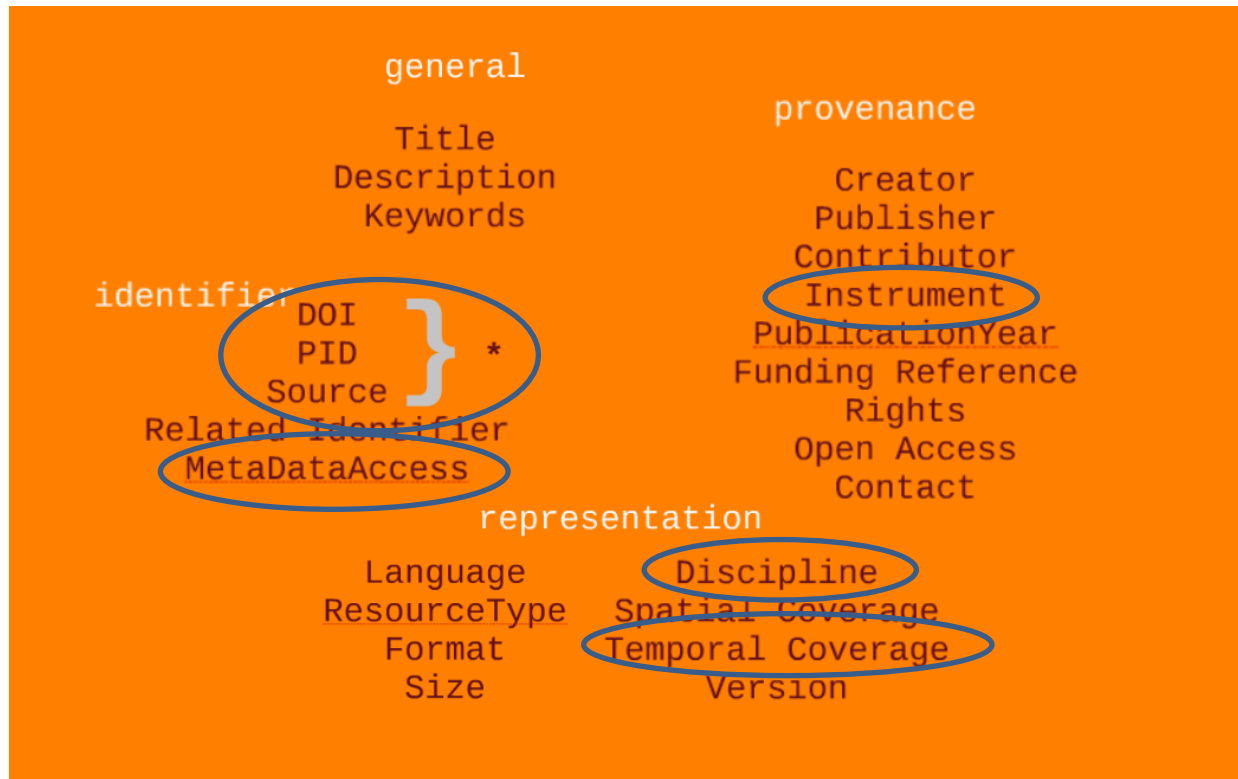
- free text search
- faceted search
e.g. spatial/temporal coverage
- narrow down results

Mapping

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



- supported Reader
 - generic: Datacite, DublinCore
 - thematic: ISO19139, FGDC
 - specific: FF (DDI and CMDI in old version)
- customised mapping
 - for each Community -> curation!



**why yet another
generic metadata
schema?**

because we
need it...

based on Datacite
additional elements

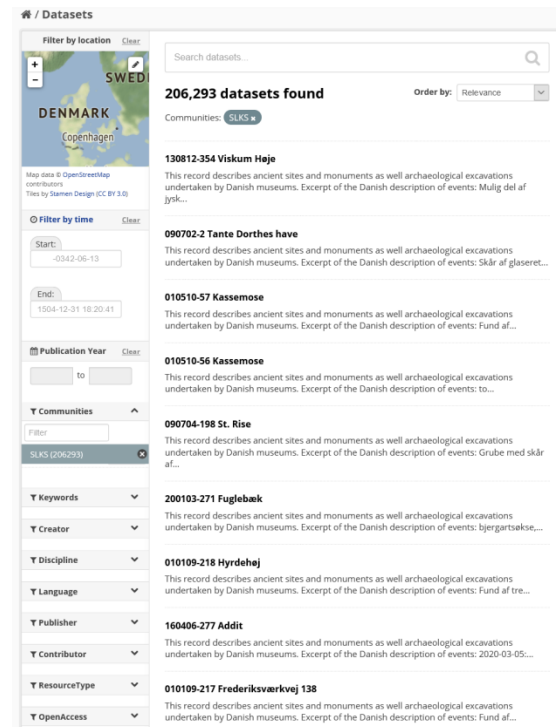
- Community
- Discipline
- Meta Data Access
- Instrument
- Temporal Coverage

generic metadata standards

B2FIND	Datacite	OpenAire
Community	[prefix]	Community
Identifier either DOI / PID / Source	only DOI	Identifier types DOI, Handle, PURL, URL, URN, ARK
Discipline	Subject	Subject
Instrument	-	-
Contact	within <contributorType = ContactPerson>	within <contributorType = ContactPerson>
Temporal Coverage	Date	Date
Spatial Coverage	GeoLocation	GeoLocation

Integration of 'nordic' archaeological resources in B2FIND

- SMR = “sites and monuments records” -> geospatial information
- records from **SLKS** (Slots- og Kulturstyrelsen), the *Danish Agency for Culture and Palaces*
- Collection from *Fund og Fortidsminder* = already existing OAI-PMH, but not standardised md prefix <fbb> and <ff>
- Issues:
 - inconsistencies , not defined as XSD, no valid namespace declaration -> difficult for software using XPATH rules
 - no maintenance by Danish Agency for Culture and Palaces
- Lessons learned:
 - always have full control over your metadata ... -> staging area University Aarhus
 - not sufficient ingestion software -> refactoring code!
- D5.1 <https://zenodo.org/record/4607189#.YGuLHD9CQ2x>



The screenshot shows the B2FIND 'Datasets' page. On the left, there's a map of Denmark with a filter by location set to 'Denmark'. Below the map, there are filters for time (Start: -0342-05-13, End: 1504-12-31 18:20:41) and publication year. The main content area shows a search bar with 'Search datasets...' and a dropdown menu for 'Communities' with 'SLKS' selected. Below this, it says '206,293 datasets found'. A list of datasets is displayed, including '130812-354 Viskum Høje', '090702-2 Tante Dorthes have', '010510-57 Kassemose', '010510-56 Kassemose', '090704-198 St. Rise', '200103-271 Fuglebak', '010109-218 Hyrdehøj', '160406-277 Addit', and '010109-217 Frederiksværkvej 138'. Each entry has a brief description of the archaeological sites and monuments.

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

Code refactoring became code rewriting...

■ **B2FIND ingestion code v3.0**

- modularised and flexible software stack
- 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

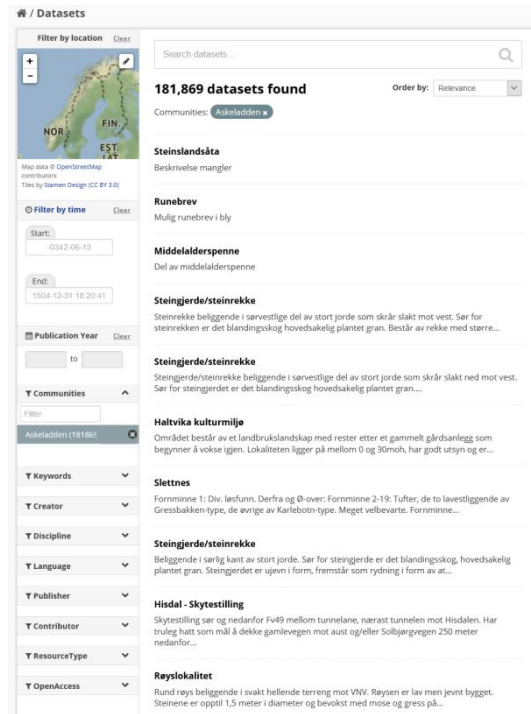


Integration of 'nordic' archaeological resources in B2FIND

- SMR = “sites and monuments records” -> geospatial information
- records from norwegian repository **Askeladden**
- harvesting via REST API (ArcGIS)
- no problem with new code!

Next steps

- Creating one 'Community' in B2FIND for 'Nordic Archaeological data' and integrating all records from SLKS and Askeladden within this group
- technically: 2 harvesting endpoints with 2 different methods (REST API and OAI-PMH) and 2 different mapfiles for specific mapping: <https://github.com/EUDAT-B2FIND/md-ingestion/tree/master/mdingestion/community>
- integrating further repos (e.g. Swedish Rock Art Research Archives, SEAD, Fornsök, Museovirasto)
- Integrating 'ontologies' (= terms related to other terms in a defined way) on search side for better precision and recall



The screenshot shows the B2FIND 'Datasets' page. On the left, there are filters for 'Filter by location' (a map of Scandinavia), 'Filter by time' (Start: -0042-06-10, End: 1504-12-31 18:20:41), 'Publication Year', 'Communities' (Askeladden 181860), 'Keywords', 'Creator', 'Discipline', 'Language', 'Publisher', 'Contributor', 'Resource type', and 'OpenAccess'. The main content area shows '181,869 datasets found' and a list of datasets including 'Steinslandsåa', 'Runebrev', 'Middelalderspenne', 'Steingjerde/steinrekke', 'Haltvika kulturmiljø', 'Slettnes', and 'Hisdal - Skytestilling'. Each entry has a brief description.

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

Integration of ExPaNDS / PaNOSC Communities in B2FIND

- on productive B2FIND
 - **ESS** URL = 'https://scicat.esss.se/openaire/oai'
SCHEMA = SchemaType.DataCite
SERVICE_TYPE = ServiceType.OAI
 - **PSI** URL = 'https://oai.datacite.org/oai'
SCHEMA = SchemaType.DataCite
SERVICE_TYPE = ServiceType.OAI
- on B2FIND testmachine
 - **ESRF** URL = 'https://icatplus.esrf.fr/oaipmh/request'
SCHEMA = SchemaType.DublinCore
SERVICE_TYPE = ServiceType.OAI
 - **ILL** URL = 'https://data.ill.fr/openaire/oai'
SCHEMA = SchemaType.DataCite
SERVICE_TYPE = ServiceType.OAI

Summary

- harvesting endpoints
 - OAI-PMH
for scicat, icat or directly
 - Datacite
- metadata standards
 - Datacite
 - DublinCore
- testcase for additional keywords
<http://eudat7-devel.dkrz.de/dataset>

Integration of ExPaNDS / PaNOSC Communities in B2FIND

Ideas / Questions

harvesting

- ❖ via OAI-PMH for (S)ICAT?
- ❖ or REST API?
- ❖ using 'search API'?

mapping

- ❖ Instrument -> how?
(DICE WP 4.2: PIDs for Instruments)
- ❖ Discipline?
(clara.science)
- ❖ additional keywords = 'PaN'?

metadata standards

- ❖ generic or specific ones?
 - <identifier> DOI/PID or else?
 - <discipline> hmmm...
 - <instrument> useful?
- ❖ DCAT reader in B2FIND
(as part of FAIR's FAIR project)

general

- ❖ Community 'group' = PaN?
- ❖ Granularity!
- ❖ integrating thesauri?

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/etc/b2find_disciplines.json

contact

EUDAT RT

<https://eudat.eu/contact-support-request>

via email B2FIND

martens@dkrz.de

fluegel@dkrz.de