IVOA Registry and EUDAT B2FIND

Harvesting the IVOA Registry for B2FIND

André Schaaff

CDS
CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG
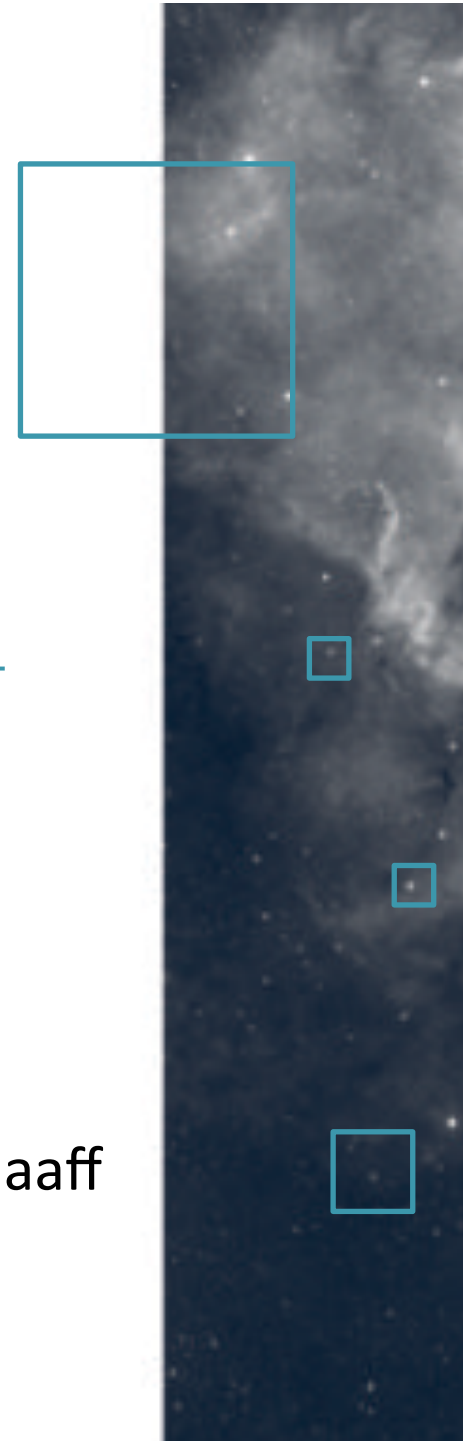
☐ EUDAT ?

« EUDAT: the collaborative Pan-European infrastructure providing research data services, training and consultancy for Researchers, Research Communities, Research Infrastructures & Data Centres »

EUDAT 1 (FP7) from October 2011 to February 2015
EUDAT 2 in the frame of Horizon 2020

EUDAT is involved in the Research Data Alliance

☐ Remark

I am not involved in the EUDAT project but i have participated to EUDAT conferences and user forums

EUDAT provides Services to the research community

☐ B2FIND ?

- B2FIND is a simple and user-friendly metadata catalogue of research data collections stored in EUDAT data centres and other repositories

- This service allows users to:
  - find collections of scientific data quickly and easily (independent from origin, discipline or community)
  - get quick overviews of available data
  - browse through collections using "standardized facets"

☐ Who benefits from it ?

- This service is primarily aimed at the researchers who want to:
  - quickly find useful data resources which can be used for their research purposes
  - build new (virtual) collections of data to address specific research questions
  - get quick overviews of available data
  - share comments on the data and metadata with other users, social networks

How does it work ?

- The metadata is harvested from the metadata providers using standard technologies (OAI-MPH but not only)

- The community itself decides which metadata is made available through B2FIND

- A sophisticated framework ensures that metadata providers are harvested regularly to always display complete and up to date information

- EUDAT provides an optimized translation from community metadata schema to standard facets in the B2FIND metadata catalogue

Some B2Find features

- Metadata is mapped onto standardized facets

- Supports faceted, geospatial and temporal metadata searches

- Allows users to search and browse datasets via keyword searches

- Results displayed in user-friendly format and listed in order of relevance

- Access to the scientific data objects is given through references provided in the metadata

☐ Faceted search

□ IVOA Registry

- IVOA Registry can be harvested through OAI-MPH

- B2Find people have harvested a small set of entries of http://registry.euro-vo.org/oai.jsp in their B2FIND test implementation

- We have worked on the mapping of the metadata

☐ Metadata mapping

Metadata field definitions

General information

The core B2FIND schema contains a number of fields as described below. In case your metadata properties do not cover or mismatch with this list, B2FIND is open to adapt and extend the B2FIND fields and the associated mapping to your specific needs.

Title (mandatory)

One line heading that names a resource, gives a description of it. A title could for example be the name of (a section of) a written resource. A title field needs to have exactly one value.

Description (recommented)

A reference to a part of the resource, for example: a table of contents, an abstract, the method employed to obtain the resource, or a single topic. Alternatively, it can be an annotation created as part of the metadata. Notes like this can also ta...

Tags

Keywords to assign ...
presented with eas ...
value of any field c ...

Provenance data

Discipline

A scientific discipline the resource is related to

Creator

The main researchers involved in producing the data, or the authors of the publication in priority order

Publisher

The name of the entity that holds, archives, publishes, prints, distributes, releases, issues or produces the resource

Maintainer (optional)

An entity that curates the data

PublicationYear

The year in which the resource was made known to the public. The field can only have one metadata field value attached to it.

Access data

Source

The URL to the a digital resource described by the metadata. The URL can be considered as a digital object identifer. The identifier should point to the resource itself, not it's metadata. The 'Source' field is mapped into from metadata fields identifying original resources like for example text or some digital audio or video format.

PID (optional)

"The PID is an alternate identifier and displayed if it exists and is not already shown in the field Source. '

DOI (optional)

"The DOI is an ...

...ghts

...y rights inf...

Formal data

Spatial Coverage

Description of geographic properties of the source described in the metadata that cannot not be termed as a location. For example: savanna. For location please use Spatial coverage. Multiple field values can be attached to this field.

Temporal Coverage

Denotation of a period in time the resource described by the metadata is associated with. The coverage is expressed by two dates: the begindate and the enddate. The dates need to be specified in UTC. Only a single value can be associated with each of these fields.
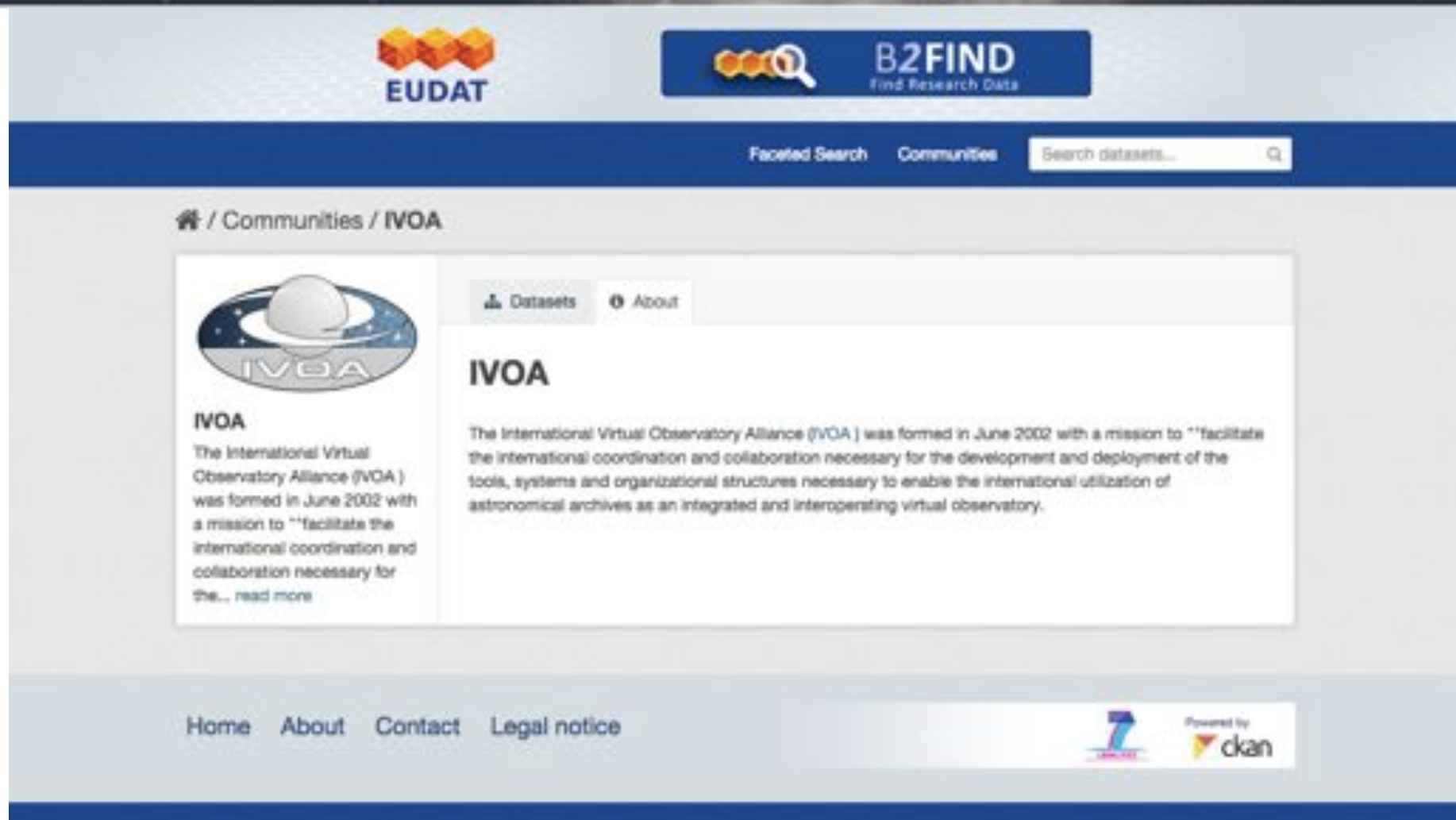
Language

Specification of the language used in the resource. For example, an English book about Japanese language, will have 'English' as the value of the language field. Multiple field values are allowed. This is mapped versus the standard language library iso639.

Format

The organisation of information according to preset specifications (computer processing) or tradition (books, physical representation of content). Multiple field values are allowed, since sometimes a metadata description applies to more than one resource.

IVOA as one of the B2FIND communities

☐ B2FIND communities

ALEPH
ALEPH was a particle physics experiment installed on the Large...
185 Datasets

B2SHARE
EUDAT offers B2SHARE - a user-friendly, secure, robust, reliable and...
223 Datasets

CESSDA
CESSDA is an umbrella organisation for the European national data archives....
529 Datasets

CLARIN
The Common Language Resources and Technology Infrastructure CLARIN project is...
137451 Datasets

DARIAH
The grand vision for the Digital Research Infrastructure for the Arts and...
2332 Datasets

DataCite
DataCite is a not-for-profit organisation formed in London on 1 December...
241180 Datasets

PaNdata
PaNdata - the Photon and Neutron data infrastructure initiative - brings...
930 Datasets

PDC
The PDC is a database of metadata and data that describes, indexes, and...
1959 Datasets

NARCIS
NARCIS provides access to scientific information, including (open access)...
175442 Datasets

GBIF
The Global Biodiversity Information Facility (GBIF) was established by...
13814 Datasets

EARLINET
EARLINET, the European Aerosol Research Lidar Network, (www.earlinet.org) is...
5 Datasets

ENES
The European Network for Earth System modelling (ENES) provides information...
830 Datasets

The European Library
The European Library comprises digital resources supplied by a prestigious...
26958 Datasets

SDL
Search Digital Libraries (SDL) is hosted by the Documentation Research and...
23903 Datasets

☐ Datasets

IVOA Registry and EUDAT B2FIND

☐ Focus

Conclusion & questions

- A way to appear with other communities

- Interest of being queried in a unique interface mixing several disciplines ?

- Live demo in Sydney

☐ Thanks

Many thanks to:

Heinrich Widmann and Hannes Thiemann

DKRZ, Germany

(Deutsches Klimarechenzentrum)

IVOA Registry and EUDAT B2FIND