

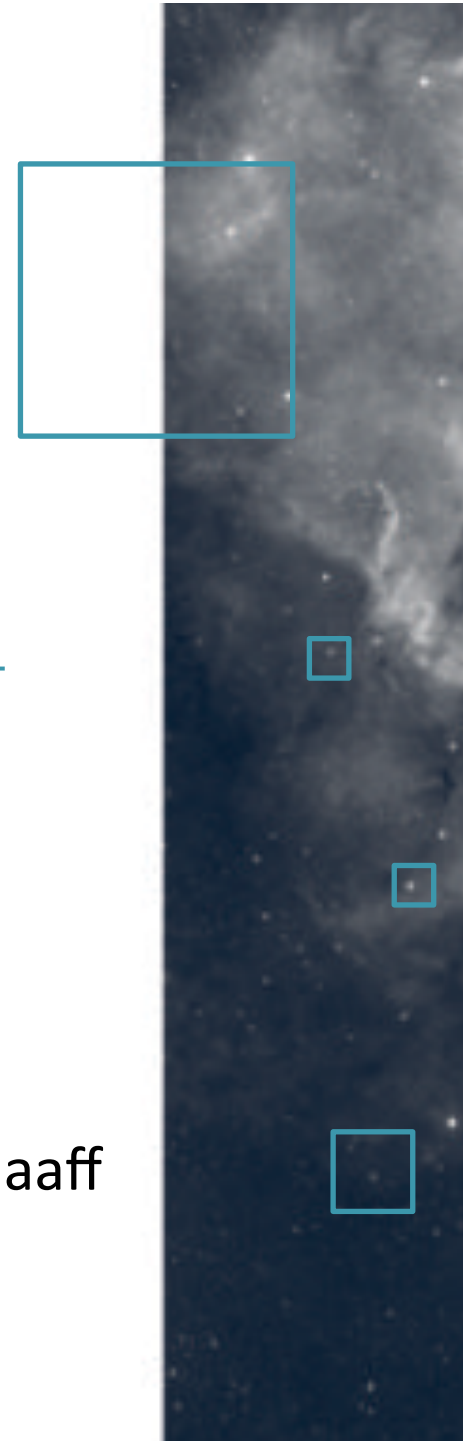
# IVOA Registry and EUDAT B2FIND

---

Harvesting the IVOA Registry for B2FIND



André Schaaff



# □ 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

Research Data Services, Expertise & Technology Solutions

f t in y p s Q

**EUDAT**

SERVICES & SUPPORT ▾ COMMUNITIES WORKING GROUPS ▾ EVENTS ▾ NEWS & PUBLICATIONS ▾ ETRAINING

EUDAT: the collaborative Pan-European infrastructure providing research data services, training and consultancy for

- Research Communities
- Research Infrastructures & Data Centres

- B2DROP**  
Sync and Exchange Research Data  
Read more | Use
- B2SHARE**  
Store and Share Research Data  
Read more | Use
- B2SAFE**  
Replicate Research Data Safely  
Read more
- B2STAGE**  
Get Data to Computation  
Read more
- B2FIND**  
Find Research Data  
Read more | Use

Use Documentation | Help Desk | Our Communities

## □ 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

The screenshot displays the EUDAT B2FIND search interface. At the top, the EUDAT logo and the B2FIND logo (Find Research Data) are visible. Below the logos, there are navigation links for 'Faceted Search' and 'Communities', along with a search bar containing the text 'Search datasets...'. The main content area is titled '/ Datasets' and features a search bar with the text 'Search datasets...'. Below the search bar, it indicates '3 datasets found' and provides an 'Order by: Relevance' dropdown menu. The search results are filtered by 'Discipline: Biology', 'Tags: Dinoflagellates', and 'Author: Pagou, Kalliopi (2008)'. The results list three datasets, all by Gotsis-Skretas, Olympia; Pagou, Kalliopi (2008):

- Microplankton abundance in the euphotic zone of the Aegean and Levantine Seas...  
Gotsis-Skretas, Olympia; Pagou, Kalliopi (2008): Microplankton abundance in the euphotic zone of the Aegean and Levantine Seas in February 1989, doi:10.1594/PANGAEA.592000
- Microplankton abundance in the euphotic zone of the Aegean Sea in October-Nov...  
Gotsis-Skretas, Olympia; Pagou, Kalliopi (2008): Microplankton abundance in the euphotic zone of the Aegean Sea in October-November 1991, doi:10.1594/PANGAEA.592040
- Microplankton abundance in the euphotic zone at Station LIA-S-GN3619850010826...  
Gotsis-Skretas, Olympia; Pagou, Kalliopi (2008): Microplankton abundance in the euphotic zone at Station LIA-S-GN3619850010826702 in the Aegean Sea, doi:10.1594/PANGAEA.592115

On the left side of the interface, there are several filter sections:

- Filter by location:** Includes a map of the world with a red box highlighting the Aegean and Levantine Seas region.
- Publication Year:** A range filter set from 1970 to 2010.
- Communities:** Lists 'OBIF (3)' and a 'Show More Communities' link.
- Tags:** Lists 'Dinoflagellata (2)' and a 'Show More Tags' link.
- Author:** A section for filtering by author.

## □ 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 (recommended)

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 be used.

### Provenance data

#### Tags

Keywords to assign presented with ease value of any field

#### 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 identifier**. 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.<sup>1</sup>

### DOI (optional)

## Formal data

### Rights

### Copyright

#### 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.

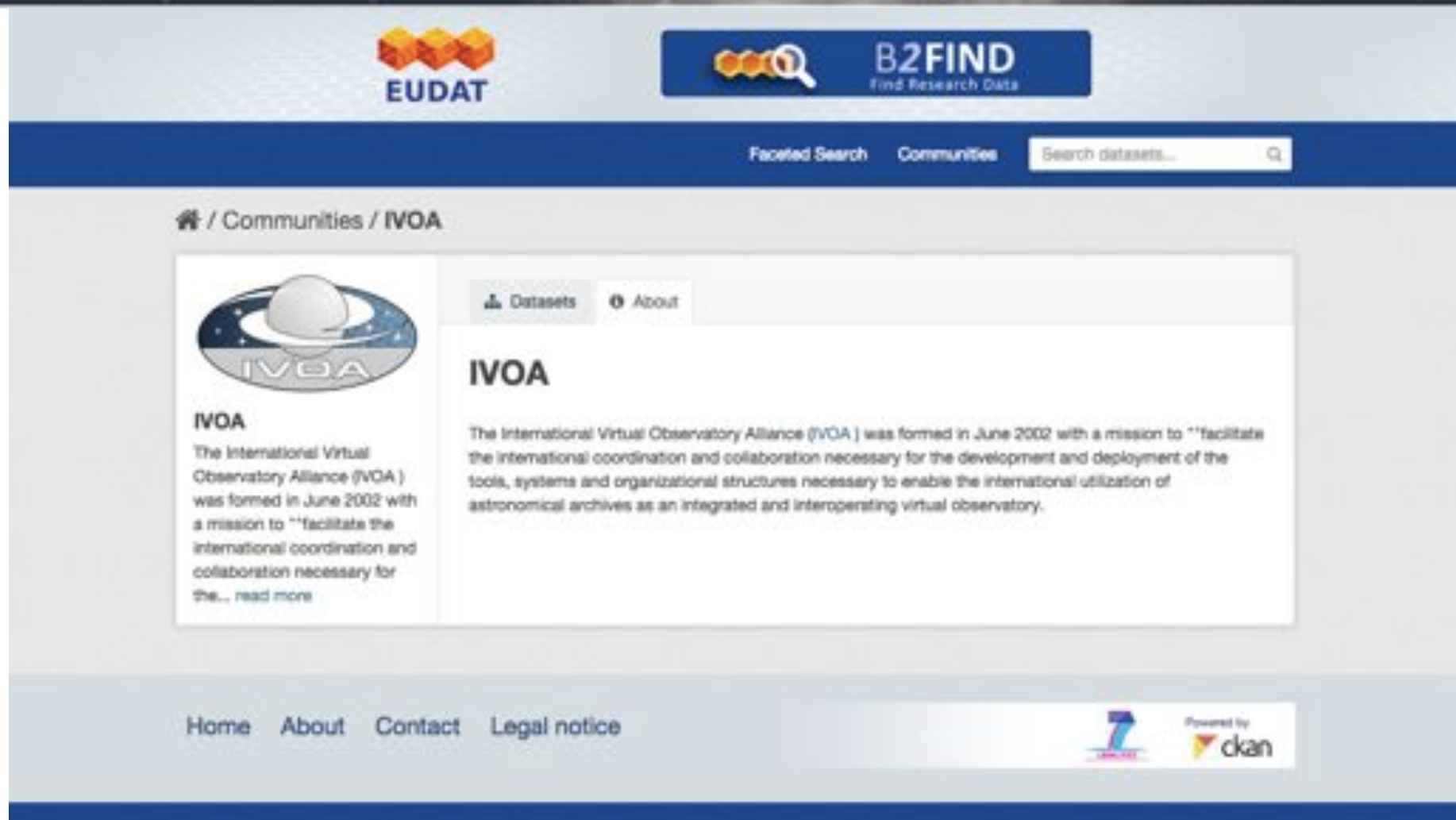
#### Language

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



The screenshot displays the B2FIND website interface. At the top, there are logos for EUDAT and B2FIND (Find Research Data). A navigation bar includes links for 'Faceted Search' and 'Communities', along with a search box labeled 'Search datasets...'. The main content area is titled '/ Communities / IVOA'. It features the IVOA logo, which depicts a globe with a ring around it. Below the logo, there are two tabs: 'Datasets' and 'About'. The 'About' tab is selected, showing the title 'IVOA' and a paragraph of text: 'The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with a mission to "facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory."'. At the bottom of the page, there are navigation links for 'Home', 'About', 'Contact', and 'Legal notice', and logos for 'Powered by' and 'ckan'.

# □ B2FIND communities

 <p><b>ALEPH</b> ALEPH was a particle physics experiment installed on the Large... <b>185 Datasets</b></p>	 <p><b>B2SHARE</b> EUDAT offers B2SHARE - a user-friendly, secure, robust, reliable and... <b>223 Datasets</b></p>	 <p><b>CESSDA</b> CESSDA is an umbrella organisation for the European national data archives... <b>529 Datasets</b></p>			
 <p><b>CLARIN</b> The Common Language Resources and Technology Infrastructure CLARIN project is... <b>137451 Datasets</b></p>	 <p><b>DARIAH</b> The grand vision for the Digital Research Infrastructure for the Arts and... <b>2332 Datasets</b></p>	 <p><b>DataCite</b> DataCite is a not-for-profit organisation formed in London on 1 December... <b>241180 Datasets</b></p>	 <p><b>PaNdata</b> PaNdata - the Photon and Neutron data infrastructure initiative - brings... <b>900 Datasets</b></p>	 <p><b>PDC</b> The PDC is a database of metadata and data that describes, indexes, and... <b>1959 Datasets</b></p>	 <p><b>NARCIS</b> NARCIS provides access to scientific information, including (open access)... <b>175442 Datasets</b></p>
 <p><b>GBIF</b> The Global Biodiversity Information Facility (GBIF) was established by... <b>13814 Datasets</b></p>	 <p><b>EARLINET</b> EARLINET, the European Aerosol Research Lidar Network, (www.earlinet.org) is... <b>5 Datasets</b></p>	 <p><b>ENES</b> The European Network for Earth System modelling (ENES) provides information... <b>800 Datasets</b></p>	 <p><b>The European Library</b> The European Library comprises digital resources supplied by a prestigious... <b>26958 Datasets</b></p>	 <p><b>SDL</b> Search Digital Libraries (SDL) is hosted by the Documentation Research and... <b>23903 Datasets</b></p>	



# □ Datasets

The screenshot displays the EUDAT B2FIND interface. At the top, the EUDAT logo (three orange cubes) and the B2FIND logo (a magnifying glass over three orange cubes) are visible. Below the logos, there are navigation links for 'Faceted Search' and 'Communities', and a search bar containing the text 'Search datasets...'. The main content area is titled '/ Communities / IVOA'. On the left, there is a sidebar for the IVOA community, featuring the IVOA logo and a description: 'The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with a mission to facilitate the international coordination and collaboration necessary for the... read more'. Below the description are sections for 'Communities' (listing IVOA with 179 datasets) and 'Tags' (listing 'star (11)', 'source catalog (11)', 'HIPASS Project Data (7)', 'Authority (5)', and 'surface brightness ... (3)'). The main content area has a sub-search bar with 'Search datasets...' and a '179 datasets found' result count. An 'Order by: Relevance' dropdown menu is also present. The first dataset listed is 'XMM-Newton SIAP Service for Pointed Observation', followed by 'Local star-forming galaxies catalog - Universidad Complutense de Madrid' and 'Theoretical Astrophysical Observatory'.

# □ Focus

Social

- Google+
- Twitter
- Facebook

Dataset Communities Activity Stream Related

## XMM-Newton SIAP Service for Pointed Observation


The European Space Agency's (ESA) X-ray Multi-Mirror Mission (XMM-Newton) was launched by an Ariane 504 on December 10th 1999. XMM-Newton is ESA's second cornerstone of the Horizon 2000 Science Programme. It carries 3 high throughput X-ray telescopes with an unprecedented effective area, and an optical monitor, the first flown on a X-ray observatory. The large collecting area and ability to make long uninterrupted exposures provide highly sensitive observations. Since Earth's atmosphere blocks out all X-rays, only a telescope in space can detect and study celestial X-ray sources. The XMM-Newton mission is helping scientists to solve a number of cosmic mysteries, ranging from the enigmatic black holes to the origins of the Universe itself. Observing time on XMM-Newton is being made available to the scientific community, applying for observational periods on a competitive basis.

X-Ray Satellite Ast...

### Additional Info


Field	Value
Source	<a href="http://registry.euro-vo.org/result.jsp?searchMethod=GetResource&amp;identifier=ivo://esavo/xmm/siap">http://registry.euro-vo.org/result.jsp?searchMethod=GetResource&amp;identifier=ivo://esavo/xmm/siap</a>
Creator	European Space Agency
Discipline	Astronomy
IVO	<a href="http://registry.euro-vo.org/result.jsp?searchMethod=GetResource&amp;identifier=ivo://esavo/xmm/siap">http://registry.euro-vo.org/result.jsp?searchMethod=GetResource&amp;identifier=ivo://esavo/xmm/siap</a>
MetadataAccess	<a href="http://registry.euro-vo.org/oai.jsp?verb=GetRecord&amp;metadataPrefix=oai_dc&amp;identifier=ivo://esavo/xmm/siap">http://registry.euro-vo.org/oai.jsp?verb=GetRecord&amp;metadataPrefix=oai_dc&amp;identifier=ivo://esavo/xmm/siap</a>
PublicationYear	2015
Publisher	European Space Agency





## □ 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)