B2FIND Integration

About

Documentation about integrating community metadata into the EUDAT metadata service B2FIND.

Modified: 09 November 2017

Synopsis

The EUDAT metadata service [B2FIND](#) provides a comprehensive joint metadata catalogue and a powerful discovery portal. Metadata are stored through EUDAT services such as B2SHARE and harvested from various research community repositories overarching a wide scope of research disciplines. Communities publishing metadata in EUDAT benefit from improved visibility and searchability of their research data in an interdisciplinary, pan-European scope.

B2FIND is open to discuss metadata publishing with interested communities and accompanies participants through the integration process. The EUDAT B2FIND team provides support by setting up the necessary data provider services on the community site, if required. The semantic mapping of the harvested metadata uses an elaborate and flexible software stack. This allows clearly formulated and easy implementation of the mapping rules according to your specific needs.

The integration process will cost the communities little effort and is described in the following in detail. More high-level information about the service is available from the dedicated [user documentation page](#). Detailed documentation about the usage of B2FIND can be found in the document [B2FIND usage](#).

How to publish metadata in B2FIND

The following two prerequisites must be fulfilled in order to publish your metadata in the B2FIND catalogue:

- Offering a service for providing and transferring the metadata.
- Defining a community-specific mapping.

In the following subsections we describe in more detail how this can be achieved.

How to provide metadata to B2FIND

The communities have to set up data servers that allow harvesting of the provided records via an internet protocol. B2FIND prefers to use the [OAI-PMH](#) protocol, as it is standardised, widely used and easy-to-use technology. We are open to other solutions for transferring the metadata from the community to the B2FIND server. If needed, B2FIND offers support for setting up an OAI server at the community site.

Metadata formats

B2FIND is open to integrating any metadata scheme - so, feel free to discuss your specific formats, schema and structure with us. Metadata formats that are already supported by B2FIND are listed in the following table.

Name	Specification	Description	Used by Communities
Dublincore	Specification: See at http://dublincore.org/specifications/ and in the following standard documents: • IETF RFC 5013 • ISO Standard 15836-2009 • NISO Standard Z39.85	The **Dublin Core Schema** is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks.The full set of Dublin Core metadata terms can be found on the Dublin Core Metadata Initiative (DCMI) website, see left.The original set of 15 classic metadata terms, known as the Dublin Core Metadata Element Set, are endorsed in the above standards documents.	DataCite, NARCIS, PanData, TheEuropeanLibrary, SDL
ISO 19115	http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53798	**ISO 19115-1:2014** defines the schema required for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services.	ENES

Name	Specification	Description	Used by Communities
MarcXML	http://www.loc.gov/standards/marcxml/	**MARC** (**MA**chine-**R**eadable **C**ataloging) standards are a set of digital formats for the description of items catalogued by libraries, such as books. It was developed by Henriette Avram at the US Library of Congress during the 1960s to create records that can be used by computers, and to share those records among libraries.	B2SHARE and ALEPH
CMDI	http://www.clarin.eu/content/component-metadata	**CMDI** (**C**omponent **M**eta**D**ata **I**nfrastructure) was initiated by CLARIN to provide a framework to describe and reuse metadata blueprints. Description building blocks ("components", which include field definitions) can be grouped into a ready-made description format (a "profile").	CLARIN
DDI	http://www.ddialliance.org	**DDI** (**D**ata **D**ocumentation **I**nitiative**)** is an effort to create an international standard for describing data from the social, behavioural, and economic sciences.	CESSDA

How to map and ingest metadata in B2FIND

Development of an adapted converting and semantic mapping is necessary before the metadata records can be uploaded into the B2FIND catalogue and made searchable in the B2FIND portal.

Metadata mapping and the B2FIND schema

The harvested "raw" metadata records are community-specific with regards to the metadata format (see above)

and to the content, i.e. the property definitions and values. As a result, a consultation needs to be held with the new community to determine how the mapping will be configured and adapted to the community-specific needs.

The core B2FIND schema contains a number of metadata definitions as described below. In case your metadata properties are not covered or do not fully match with this list, B2FIND is open to adapt and extend the B2FIND fields and the associated mapping to your specific needs. The only mandatory item is 'Title'. All other fields are optional but recommended, as we intend to use most of them in order to achieve the best possible coverage.

Title	A name or title by which a resource is known.
Description	Additional information describing the content of the resource. Could be an abstract, a summary or a Table of Contents.
Tags	A subject, keyword, classification code, or key phrase describing the content.

Source	The Source is an identifier, therefore a unique string that identifies the resource. It may link to the data resource itself or to a landing page that curates the data.
PID	The PID is an alternate identifier.
DOI	The DOI is an alternate identifier.

Community	Research communities that provide research data to EUDAT. Could be an aggregator as well.
Discipline	A scientific discipline the resource originates from. A closed vocabulary based on a Wikipedia-classification is used.
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.
Publication Year	The year in which the resource was or will be made publicly available.

Language	The primary language of the resource. Codes are mapped to long names according to ISO 639.
Temporal Coverage	Period of time the research data resource is related to. Could be a date format or plain text or both.
Spatial Coverage	A geolocation the research data resource is related to. Could be geographic coordinates of the Earth's surface (e.g. longitude/latitude) or denomination of places.
Format	Technical format of the resource.

Contact	Any contact information for this resource.
MetadataAccess	The [OAI-PMH GetRecord](#) request.

Metadata upload to the B2FIND portal

After the iterated process of adaption and review of the mapping reaches an agreed state, an initial upload of the mapped records is performed. A regular and incremental ingestion workflow assures real-time synchronisation between the data pool harvested from the communities and the datasets provided and made searchable in the [B2FIND portal](#).

Support

You can find B2FIND *training presentations* on the [EUDAT website](#).

You can also find *hands-on training material* on B2FIND on our [github repository](#).; note in particular sections 00-04.

Support for B2FIND is available via the EUDAT ticketing system through the [webform](#).

If you have *comments* on this page, please submit them though the [EUDAT ticketing system](#).

Document Data

DocVersion: 1.2.1

Authors:

Heinrich Widmann, widmann@dkrz.de

Hannes Thiemann, thiemann@dkrz.de

Editors:

Hans van Tiggelen, hans.vanpiggelen@surfsara.nl

Kostas Kavoussanakis, k.kavoussanakis@epcc.ed.ac.uk

Sara Ramezani, sara.ramezani@surfsara.nl

Read more