Frequently Asked Questions

B2SERVICES

- **Are the decisions in EUDAT services driven by the metadata available from the EUDAT research communities?**
- **Can the EUDAT B2 services be installed as local instances? Only some of them? How to connect them? What technology options are there?**
- **Are there limitations to the file size and number of files for the B2 Services?**
- **Why should I have the data in different places: B2SHARE, B2DROP, ...?**

ONTOLOGIES & SEMANTIC WEB

- **What is the difference between a controlled vocabulary and an ontology?**
- **How do I write upper ontologies for semantically linked data?**
- **Where do I find existing ontologies - good resources for finding this information?**

DATA MANAGEMENT PLANNING

- **What are the main challenges in writing a Data Management Plan?**
- **What is data curation exactly? Does it cover the full data life cycle or is it limited to the data creation phase & storage?**

B2 Services

Are the decisions in EUDAT services driven by the metadata available from the EUDAT research communities?

No. All of the services are either agnostic to metadata or can support community specific metadata schema's.

- B2DROP service has no specific support for handling metadata, although if a user chooses to deposit a separate metadata file they can, but B2DROP does not maintain a link between the file and metadata.
- B2SHARE can be configured to support community-specific schema's, and already has support for a number of existing communities. To see how to add metadata to records for existing schemas see B2SHARE Usage Documentation. To create a new community and add a new schema (or modify an existing one) visit the B2SHARE installation training pages.
- The current version of B2SAFE, like B2DROP, does not provide a means of associating metadata and data objects, or any way of extracting metadata from containers (such as tar files). However, this support is being developed currently and is in the alpha stage, and the intention is not to impose metadata standards on communities but support arbitrary schemas. At the current time the information on B2SAFE metadata handling can be found here.
- B2FIND can harvest from any existing metadata schema (subject to certain minimal fields being available), but may not display all of the harvested fields. However, all harvested fields are searchable. For more details on how to integrate your own metadata into B2FIND visit the B2FIND Integration documentation.

The remaining services have no interaction with metadata schemas, although B2NOTE can be considered as a means of providing arbitrary metadata..

Can the EUDAT B2 services be installed as local instances? Only some of them? How to connect them? What technology options are there?

Some of the B2 services can be installed locally with differing levels of complexity. .

- **B2DROP** is based on the common [nextCloud](#) technology with some customisations (such as the UI and the ability to push data into B2SHARE). This is relatively easy to install locally, but local administrators will need additional support to link to either a local B2SHARE instance or the EUDAT provided one. The [B2DROP github](#) repository contains both puppet manifests and docker images to support local installation.
- **B2SHARE** is currently available as a docker image for local installation on the [B2SHARE github](#) repository.
- **B2SAFE** represents a series of extensions to [iRODS](#). These extensions are available from the [B2SAFE github](#) repository. Full instructions for local installation are available in the B2SAFE Deployment section of the [EUDAT User Documentation](#). If you wish to provide additional rules for local reasons, currently EUDAT does not supply official support but you may be able to get assistance from the iRODS community
- **B2STAGE** - Deployment instructions can be found in the [B2STAGE Site Administrators](#) documentation.
- **B2FIND**, like B2DROP, requires you to install the [CKAN](#) discovery portal and add B2FIND extensions (available from [B2FIND github](#) repository). However, there is no official support for local installations.
- The **B2HANDLE** python library can be used independently from having any agreements with EUDAT. Full details are available in the [B2HANDLE for communities](#) documentation. However, to create Handles yourself you need a prefix and access to a local Handle server. You can become part of a persistent identifier federation such as EPIC or DOI and setup your own local handle server, but you can also get access to an EUDAT handle server by having agreements with a data centre running Handle servers.
- **B2NOTE** - This is currently not a full production service and we do not recommend trying a self installation.
- **B2ACCESS** - Currently there is no support for local instances of B2ACCESS.

Are there limitations to the file size and number of files for the B2 Services?
In principle there is no limit to the number of files that can be added to each service (although there is a finite amount of storage currently associated with each service). In terms of file sizes, there are limitations for some services as described below.

- B2DROP: 2GB per file, with a maximum of 20GB per user.
- B2SHARE: Currently the maximum file size is 10 GB and the maximum size for a record is 20GB. The current limits are given [here](#).
- B2SAFE: unlimited, but very large file transfers are subject to network stability of course.

Why should I have the data in different places: B2SHARE, B2DROP, …?
The different services are tailored towards different types of data and different functionalities (enabling different states of the data life cycle as found in the [training guides](#)) pertaining to level of sharing, size of data and team, availability, etc.

- **B2DROP** - for small size data, no metadata, sharing in smaller teams and external peers (not having an account in B2DROP), for data that is actively used, edited and used by several peers
- **B2SHARE** - for small to medium size data, rich metadata possibilities, data publishing (get citations and credit for your data). Single files in the publication are addressable and downloadable by PIDs, for data that supports a publication, data that will not change in the (near) future
- **B2SAFE** - for small to large size data, multiple copies not only in a system but even across geographical locations, automatic and regular data replication to several sites (bringing data closer to other compute services and for safe keeping of data), automatic and regular integrity checks between replicas, in combination with B2STAGE-gridFTP fast data transfers, for data that is employed in computations, kept as fall-back points (backup) or data that is still changing but needed in different places

While there is no need to have the same data in each service, in some cases it may be useful. For example, you can move version 1 of a data set from B2DROP to B2SHARE to publicise it, but you may want to keep the copy in

B2DROP to allow for additional work on version 2 of the data set. Note, however, that currently data published in B2SHARE is not deletable, while both B2DROP and B2SAFE allow you to delete your data from the system if it is no longer required.

Ontologies and Semantic Web

!!! Note that EUDAT can not directly support communities with developing ontologies. For more discussions on this topic please contact the leaders of the EUDAT Semantics working group or Vocabulary Services Interest Group of the Research Data Alliance.

What is the difference between a controlled vocabulary and an ontology?
There are no real differences between controlled vocabularies and ontologies. An ontology is a conceptual model of the world representing the interdependencies between the concepts. A controlled vocabulary is a simple form of an ontology often representing the hierarchical relation between concepts (i.e. superclass/subclass). Ontologies can be more complex and include a model based on first-order logic that will add associative and dependency relations between classes. We are then talking about formal ontologies.

How do I write upper ontologies for semantically linked data?
An upper ontology is conceptual model of the world that serves as a basis for integrating more domain specific ontologies. It is used for interoperability between ontologies. Several upper ontologies exist such as CIDOC-CRM, the Basic Formal Ontology, SUMO, DOLCE. You can actually look at the list of existing upper ontologies on the dedicated wikipedia article (https://en.wikipedia.org/wiki/Upper_ontology). Regarding the format, upper ontologies can be serialized using OWL which will then support first-order logic if needed. Please keep in mind before adding yet another upper ontology that a good practice is to reuse what exist already.

Where do I find existing ontologies - good resources for finding this information?
That is a very good question to which it is hard to answer at the moment. In the biomedical domains, several ontology repository exist such as Bioportal, EBI OLS, Ontobee. In other domains, you can find vocabulary repositories but unless you are in field they can be hard to find. For Earth Science, you have for instance ESIP Ontology portal or Agroportal for agriculture. This subject has been the topic of a dedicated workshop organized in April 2016. The work started in this workshop is now going to be continued within the RDA Vocabulary and Semantic Service Interest Group.

Data Management Planning

!!! Note that EUDAT does not provide support to assist communities writing their data management plans but works with international partners such as the Digital Curation Centre and DANS and the OpenAIRE project on providing services to support data management.

What are the main challenges in writing a Data Management Plan?

- Making writing the DMP a team effort to get team commitment to DM during the project.
- Linking it sensibly to the researchers' workflow and standards, to avoid mere administration.
- Remembering that ultimately data management is for re-users: what will they need?
- Making it detailed enough.

What is data curation exactly? Does it cover the full data life cycle or is it limited to the data creation phase & storage?
There is no strict definition of the difference between curation and preservation of digital data and the two are

often used interchangeably, which can lead to confusion. Some communities do not distinguish between the two while others look at them differently. For instance, data preservation may be considered as preserving the actual bits you record with no corruption over time, while curation is ensuring that the data you write now will still be readable in 20 years time (imagine trying to read an excel spreadsheet from excel 2.0 into the current edition). Opinions differ on whether the term "data curation" covers the whole lifecycle and there is no one right answer to this. Data Management certainly applies to the whole life cycle of research data.

Read more