

## D7.1 Annex 1: A comparison of infrastructures for publishing research data

	B2Share	DataverseNL	Figshare	Dryad	Zenodo	Mendeley
<b>Who owns the data?</b>	Ownership remains with depositor; B2Share is executor	Ownership is not transferred	Ownership is not transferred	As the CC0 licence is being used, there is no formal ownership of the data.	Ownership is not transferred	Ownership is not transferred
<b>Licensing of deposits</b>	License indicated by depositor in free-text metadata field. EUDAT suggestion: CC BY 4.0; ODbL v1.0; Open Data Commons Attr. License v1.0.	Terms of use available when uploading and downloading. Additional terms may be specified for individual dataverses and studies.	CC0 or CC BY	All submissions under CC0 license	Metadata under CC0 license. Licensing of data files to be decided by depositor.	CC BY 4.0
<b>Responsible for checking deposits</b>	The depositors themselves	Dataverse administrators/ curators control release of study.	The depositors themselves	Dryad checks deposit before inclusion. Depositors are responsible for content.	Zenodo checks deposit before inclusion.	Data input is moderated, not curated. Datasets must be scientific in nature.
<b>What is the owner allowed to do?</b>	Make data public/restricted (in general terms, not for specific individuals)	Release/Discard. Assign curators/contributors. Restrict data, e.g. by granting access to specific user roles only.	Make data public/ private/ draft. Add, delete, replace files.	Upload additional data files	Make data public, restricted or put under embargo (in general terms, not for specific individuals)	Deposit additional data files; request dataset removal from administrator; put a published dataset under embargo.
<b>Differentiation of roles</b>	Depositors and (community) administrators, the latter with the authority to delete datasets	Administrator (Institute), Curator (Dv), Contributor (Dv), Network Administrator (DANS back office)	None	Contributor and curator, the latter with the authority to delete datasets	None	Contributor and administrator, the latter with the authority to delete datasets
<b>Who is allowed to upload?</b>	Users registered with B2Share	Users registered via SURFconext. Basically, it is the institute who assigns the right to upload.	Users registered with FigShare	Registered users	Users registered with Zenodo. Uploads may be accepted/rejected before inclusion into community collections.	Registered users
<b>Batch uploads possible?</b>	Yes, also through drag & drop	Yes, from Version 3.6.2	Yes, through Figshare API	No	Yes, through drag & drop and via Dropbox	No
<b>Versioning</b>	Not yet possible (expected in B2Share 2, autumn 2016)	Yes, post-release at study level. Adapting metadata also generates new version which has to be released.	Yes, with public studies, not with private ones.	Yes, at the study and file levels	Yes, at data file level, not at metadata level	Yes. Editing of an already published dataset will lead to a new draft version that may be published as version 2 of the dataset. All versions of any given dataset (if any) can be viewed by clicking the links in the Version History.
<b>Removal of studies or files?</b>	Yes, but these operations should be rare. Only the service administrator or the community administrator are allowed to change a public object, for extreme cases like dealing with abuse or because of legal concerns.	Only DANS network administrator is allowed to remove files and versions. Dataverse administrator can de-access studies.	Files and studies can be removed provided they are not public. Replacing files is always possible.	Files can only be removed by Dryad curators.	Studies can be withdrawn from public access, but data files and metadata are kept.	You can delete a dataset until it has been published. After that, you will need to request Mendeley to delete the dataset.
<b>Data file types and sizes</b>	Unlimited number of files. Maximum file size is 2GB. No data file type restrictions. Custom players for data types (text, audio, video) are available.	Unlimited number of files. Maximum file size is 2GB. No data file type restrictions.	Unlimited number of files. Maximum file size is 5GB, with 20GB free private storage. Unlimited public storage. No data file type restrictions.	Max. study size is 10GB. No data file type restrictions.	Unlimited number of files. Maximum file size is 2GB. No data file type restrictions.	Unlimited number of files. Any file type can be uploaded. Preview capabilities for some file types (PDF, text, etc.) are available.
<b>Persistent Identifier?</b>	Yes, Handle (prefix/suffix) at study level. Resolver at <a href="http://hdl.handle.net/">http://hdl.handle.net/</a>	Yes, Handle (prefix/suffix) at study level. Resolver at <a href="http://hdl.handle.net/">http://hdl.handle.net/</a>	Yes, at dataset or file level. A DOI is assigned to each dataset.	Yes, at the study and file levels. Also indicates different versions. A DOI is assigned to each dataset and each data file.	Yes, at the study and file levels. Also indicates different versions. A DOI is assigned to each dataset.	A DOI is assigned to each version of a public dataset.

	B2Share	DataverseNL	Figshare	Dryad	Zenodo	Mendeley
<b>Metadata format</b>	None	DDI / Dublin Core	None	Dublin Core DCAM	MARC, exported to MARCXML, Dublin Core and DataCite Metadata Schema (follows OpenAIRE Guidelines)	None
<b>Metadata Study Deposit Interface</b>	One metadata deposit interface per collection.	At study level, exclusion of specific metadata fields in deposit interface is possible.	Very limited metadata.	Also depends on journal requirements of associated paper.	None	Yes
<b>Metadata level</b>	Study level	Study level up to and including file level (free text only, in Description field)	Dataset or file level	Study and file levels	Dataset level only	Study level
<b>File organisation</b>	8 standard collections: Generic, Ontology, Hydrometeorology, Linguistics, Biodiversity, Herbarium, RDA, Biomedical Research. Study -> Files	Dataverse -> Collections -> Studies -> Files. Collections can be nested.	Project -> Fileset -> Files. Files can be stored in a Fileset.	Publication -> Files	Uploads to customised community collections	Related files are logically linked to dataset.
<b>Archiving term and sustainability</b>	Indefinite period	In practice the period is indefinite, but repository is positioned as mid-term solution (5-10 years).	Indefinite period	Indefinite period. Sustained (common formats are actively updated).	At least 20 years	Indefinite period
<b>Embargo</b>	The depositor may specify an embargo date when uploading data. The metadata will be accessible right after the upload, but the access to the actual data will be restricted until the embargo date.	Embargo not enabled; data release yes/no is possible.	Embargo is possible at the dataset and data file levels.	Embargo until publication of paper, or max. 1 year (files only).	User-driven embargo is possible.	When publishing a dataset, a user may choose to put the published dataset under embargo until a specified date (for example, to make it available at the same time as an associated article).
<b>Confidentiality</b>	Stored at CsC Kajaani data centre (Finland), which is ISO 27001:2005 certified.	Stored at Vancis (SURFsara): ISO 27001; ISO/IEC27001; NICTIZ certified. DANS does not actively monitor checksum changes.	CLOCKSS's network of redundant archive nodes (for accessibility); Amazon S3 security.	CLOCKSS's network of redundant archive nodes (for accessibility); security measures.	Stored in cloud infrastructure. Both data files and metadata are kept in multiple online replicas. Backed up to tape every night. Files' MD5 checksums are regularly checked.	Stored on Amazon S3 servers in Germany. In partnership with DANS, all data are regularly backed up for sustained availability.
<b>Continuity</b>	Data will remain archived and accessible for at least 2 years after B2S service should discontinue.	No promises ('As is')	No promises ('As is')	No promises ('As is')	Upon discontinuation, all repository content will be transferred to another suitable repository.	Discontinuation risk has been transferred to DANS.
<b>Cost</b>	None (for European scientists and researchers)	Institute pays fixed fee + usage cost. Depositor charges are up to institute.	None for up to 1GB private storage of unpublished content, free basic service.	\$80 per data package, or via journal/institute. Additional fee applies from 10GB.	None	None
<b>Openness of data</b>	Data are restricted or open access. Metadata are always open access.	Access rights possible at the dataverse, study and datafile levels. In principal, metadata are open access after release, but they may remain hidden upon request.	Data are public or private.	Data and metadata are open access after publication of the associated paper, except in case of an embargo.	Data may have open, closed or restricted access (i.e. the depositor decides if access will be permitted).	Data and metadata are open access after publication of the associated paper, except in case of an embargo.
<b>Draft version editing</b>	Not yet possible (expected in B2Share 2, autumn 2016)	Yes	Yes	Yes. Adding files is possible. File deletion only by Dryad curators.	No	Any new dataset starts off in draft mode. When it is published, it changes to public version 1.
<b>Scope</b>	Aimed at European researchers, plus specific communities	Aimed at users within DataverseNL	Aimed at researchers worldwide	Aimed at researchers worldwide	Aimed at researchers worldwide, plus specific communities	Aimed at researchers worldwide

	<b>B2Share</b>	<b>DataverseNL</b>	<b>Figshare</b>	<b>Dryad</b>	<b>Zenodo</b>	<b>Mendeley</b>
<b>Findability</b>	Public search field functionality. Harvested by B2FIND.	Public search field functionality. Supports OAI-PMH.	Public search field functionality.	Public search field functionality. Published content available for replication and indexing by third-party partners.	Public search field functionality. Harvesting of all content via OAI-PMH protocol.	List of published datasets can be browsed on data.mendeley.com. In the future, keyword search of dataset contents or metadata, and browsing by subject and other attributes will be possible.
<b>Downloading</b>	Data citations directly exportable to bibliographic software. Unregistered users can only access, not download, open data.	Permissions include Guestbook (note who downloads your data) and Email alerts for restricted access requests. Unregistered users can download open data.	Anyone can download.	Download statistics available. Data citations directly exportable to bibliographic software.	Altmetrics statistics available.	Depending on the terms of the licence you selected for your data, others may be able to publish them elsewhere, provided they give you attribution and cite the licence. Full control over who can view and download your research data.
<b>Interaction with other tools</b>	Potentially, interaction with other EUdat tools and the iRODs REST API is available.	Potentially, yes (open source). Interaction API, library for addressing DVN instances and other clients available.	Open source, many existing interactions. API available.	Interaction with journals. API under development.	Uploads directly from Dropbox. GitHub integration.	An API is available. Reference documentation: <a href="http://dev.mendeley.com/methods/">http://dev.mendeley.com/methods/</a> Interaction with journals will be made possible.
<b>Ability for users to rate datasets</b>	Yes, via 'Rate this document' option.	Yes, via study's 'Comment' tab.	Yes, via study's 'Comment' field.	No	Yes, via study's 'Comment' field.	No

Adapted from: 'Vergelijking infrastructuren voor het publiceren van onderzoeksdata' by Tessa Pronk and Kees van Eijden, 21-05-2014.