Assessment Information

Repository:	DataverseNO
Website:	https://dataverse.no
Certification Date:	24 March 2020

This repository is owned by: UiT The Arctic University of Norway

DataverseNO

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

National repository system; including governmental

Comments

DataverseNO [1] is a Norwegian national, generic repository for open research data. DataverseNO is not a separate corporate body, but is owned by, and part of, UiT The Arctic University of Norway. DataverseNO is operated by the IT Department and the University Library at UiT The Arctic University of Norway. See also re3data [2]. The repository is built on the open source application Dataverse, developed mainly at Harvard University [3]. DataverseNO is mentioned as one out of five national, generic research data services in the national policy for research data management in Norway [4]. DataverseNO accepts submissions from researchers primarily from Norwegian research institutions. These datasets are grouped into collections and sub-collections. Such collections are a way of grouping and visualizing datasets within the DataverseNO repository. DataverseNO is thus one single repository containing multiple collections, and not an aggregation of independent collections.

Norwegian research institutions can use the DataverseNO repository as partners. Each partner institution is assigned their own institutional collection within the DataverseNO repository. The division of responsibilities between UiT The Arctic University of Norway (owner of DataverseNO) and the DataverseNO partner institutions is regulated in partner agreements. Each partner institution is responsible for the stewardship of the data deposited into their institutional collection within DataverseNO according to the DataverseNO policies and guidelines. See also section on partner agreements below and R5.

A DataverseNO partner institution may also establish collections that target user group(s) not limited to the researchers at their institution. Such collections are here called special collections. The scope of special collections may be thematic, project-based, subject-based or other. TROLLing – The Tromsø Repository of Language and Linguistics [5] is a thematic, and currently the only special collection in DataverseNO. All collections within DataverseNO are at the full responsibility of the DataverseNO partner institution for whom the collection was established; in the case of TROLLing this is UiT The Arctic University of Norway. See section on partner agreements below.

Researchers who are associated with Norwegian research institutions that are not partners of DataverseNO or who are not in the user group of any special collection of DataverseNO can archive their data in the top-level collection of DataverseNO. These data are curated by Research Data Service staff at UiT The Arctic University of Norway.

The organization of DataverseNO is described in the section Organization of DataverseNO [6] of the About page on the

DataverseNO info site, and is discussed in detail in section R5.

All policies, governance and steering documents, and guidelines for all aspects of the DataverseNO repository apply to the entire DataverseNO repository including all collections. This present CoreTrustSeal application covers the entire DataverseNO repository, including technology, people, procedures, and stewardship.

In order to ensure the full compliance of all DataverseNO policies and guidelines in all their aspects throughout the entire DataverseNO repository, DataverseNO signs two different agreements with DataverseNO partner institutions:

- A partner agreement with DataverseNO for institutions that want to be assigned collection(s) within the DataverseNO repository. The agreement regulates roles and responsibilities between UiT The Arctic University of Norway (owner of DataverseNO) and the partner institution for the collection.
- A data processor agreement between UiT The Arctic University of Norway (owner of DataverseNO, data processor) and the DataverseNO partner institution (data controller). The agreement regulates the processing of personal data carried out by the data processor on behalf of the data controller in connection with the use of DataverseNO. This agreement applies to both partner institutions as well as non-partner institutions with individual researchers who archive their data in the top-level collection of DataverseNO.

These documents are available upon request.

Whenever a detailed account is not necessary, the term DataverseNO is used in this application to cover both the owner institution and all/any responsible partner institution(s). In this context, ownership means that DataverseNO is part of UiT The Arctic University of Norway and not its own corporate body.

References:
[1] https://site.uit.no/dataverseno/about/
[2] re3data.org: https://www.re3data.org/repository/r3d100012538 and https://www.re3data.org/repository/r3d100011623
[3] https://dataverse.org and https://github.com/IQSS/dataverse
[4] National policy for research data management in Norway (12/2017) –
https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf (p. 20, Norwegian only; English translation below)
[5] TROLLing – The Tromsø Repository of Language and Linguistics: https://info.trolling.uit.no
[6] Organization of DataverseNO: https://site.uit.no/dataverseno/about/#organization-of-dataverseno

English translation of National policy for research data management in Norway:
"There are five data archives/infrastructures that can be termed generic, i.e. they offer services across most areas of expertise. UNINETT Sigma2 AS has established the National e-Infrastructure for Research Data (NIRD), which offers services and capacity for all disciplines that require access to advanced large-scale resources for storing, processing and

publishing research data or searches in digital databases and collections. The Norwegian Center for Research Data (NSD) is setting up the Norwegian Open Research Data Infrastructure (NORDi), a new solution for uploading, preserving and sharing research data, which will support open access to and reuse of data from social sciences and humanities research and research in medicine, health, climate and environment. Services for Sensitive Data (TSD) at the University of Oslo (UiO) provide a full set of services for analysis, processing and storage, in a secure environment. In addition to UiO, the TSD services are also used by several other national research institutes. UiT Open Research Data is a generic infrastructure service for researchers at UiT, which additionally offers the DataverseNO service to other Norwegian research institutions that want an institutional repository for research data. The service is also open to individual researchers from Norwegian institutions who need an open archive for archiving, publishing and citing their own research data, specifically to provide an offer that meets the requirements of journals that background data should be available. Partner institutions also get access to training, support for super users and guidance/manual for curation."

Brief Description of the Repository's Designated Community.

Since the DataverseNO repository provides free and open access to its collections, the Designated Community of the repository consists of both data contributors and data users. Data users include primarily researchers and research institutions, but also any other stakeholders in society reliant on access to knowledge, e.g. journalists, teachers, industry as well as the greater public. The interaction between data users and the repository happens primarily through direct contact with the contact person(s) for each dataset (see R11), and through the general contact information provided for each collection.

The term Designated Community is used here to describe the different user groups that in addition to being data users also are data contributors to the repository. As described in the section DESIGNATED COMMUNITY [1] of the About page on the DataverseNO info site, these user groups fall into three main categories:

1) researchers from Norwegian research institutions that are partners of DataverseNO
2) researchers working within the scope of any special collection within the DataverseNO repository
3) researchers from Norwegian research institutions that are not partner of DataverseNO.

Although a single researcher may belong to more than one of these user groups, each user group relates to their dedicated collection within DataverseNO, and each collection is organized and managed in a way that ensures that the needs of the user group are met to the largest possible extent.

The DataverseNO policies and guidelines are common for all collections within DataverseNO and describe the scope, knowledge base, and methodologies – as well as the curation needs – of the Designated Community targeted by DataverseNO. The partner agreement regulates the responsibility of the partner institution to understand and comply with these policies and guidelines, as well as the responsibility of UiT The Arctic University of Norway (owner of DataverseNO) to provide necessary training for the partner institutions.

1) First type of user groups
Researchers from DataverseNO partner institutions include employees, students and other affiliates of Norwegian research institutions that have signed a partner agreement with DataverseNO. Currently, there are nine partner institutions in DataverseNO (including UiT as the owner of DataverseNO), and all of them are Norwegian universities producing research within virtually all major scholarly disciplines.

2) Second type of user groups
A DataverseNO partner institution may establish special collections as described above. Special collections cover scholarly disciplines that are offered at the DataverseNO partner institutions and, as a main rule, they are therefore management and curated by Research Data Service staff at the institution responsible for the collection. A special collection is usually established on request from a user community, and is managed and curated in close dialog with the involved user community.

Currently, TROLLing (The Tromsø Repository of Language and Linguistics) is the only special collection in DataverseNO [2]. TROLLing is under the responsibility of, and is managed and operated, by UiT The Arctic University. TROLLing accepts open research data from linguists worldwide.

3) Third type of user groups
In addition to the two user groups above, DataverseNO offers their services to researchers from Norwegian research institutions that are not partnering with DataverseNO. Data from this third user group of DataverseNO are published in the top-level collection of the repository, and they are curated by Research Data Service staff at UiT The Arctic University of Norway (owner of DataverseNO). These data may come from any subject represented at any Norwegian research institution that is not partner of DataverseNO. As mentioned earlier, UiT The Arctic University of Norway – as many of the other DataverseNO partner institutions – produce research within virtually all major scholarly disciplines. Research Data Service staff from UiT are thus very likely to cover all the potential subjects represented by researchers from this third user group. In the unlikely case where Research Data Service staff at UiT are not sufficiently familiar with the subject represented by a dataset deposited into the DataverseNO top-level collection they discuss the dataset with Research Data Service staff from other DataverseNO partner institutions and the research community at the home institution of the data author before curating the dataset.

The majority of the Norwegian universities and university colleges are already partners of DataverseNO. We therefore

emphasize that the third user group of DataverseNO currently constitutes – and is estimated to constitute also in the future – only a small part of the Designated Community of DataverseNO. All Norwegian research institutions have recognized the importance of offering their researchers (a) reliable service(s) where their research data can be curated and published according to institutional, national, and international standards and best practice recommendations. Therefore, it is – at least in a Norwegian context – highly unlikely that there will be published many datasets in DataverseNO from (a) researcher(s) from a non-partner institution without the management of the non-partner institution deciding to become a partner of DataverseNO and get their own institutional collection within the repository.

All user groups of the Designated Community of DataverseNO have in common that the first-line services offered by, and the communication with, the repository are channeled through the curator(s) of the applicable collection. For a description of the communication between the Designated Community and DataverseNO, see the section DESIGNATED COMMUNITY [1] of the About page on the DataverseNO info site.

Finally, it must be stressed that the mission of DataverseNO is to be a national GENERIC repository for open research data (see R1). Despite its generic mission, DataverseNO strives to provide subject-specific expertise as far as possible; see R6, R8, and R11. This is why, as a main rule, data deposited into institutional collections or into the top-level collection of DataverseNO are curated by Research Data Service staff who are subject specialists in addition to be trained in research data management. Special collections of DataverseNO are without exception managed and curated by permanent Research Data Service staff who are specialists within the subject at stake.

References:
[1] DataverseNO Designated Community: https://site.uit.no/dataverseno/about/#designated-community
[2] TROLLing: https://info.trolling.uit.no

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Comments

Datasets deposited into DataverseNO are reviewed/curated by Research Data Service staff before they are published. Research Data Service staff are mainly library staff working at DataverseNO partner institutions and having post-graduate level expertise within the different subjects represented by the deposited data. In addition, responsible Research Data Service staff have in-depth expertise in FAIR research data management (RDM). Typically, Research Data Service staff are (Senior) Research Librarians / Subject Librarian, but also other research support staff specialized in RDM may review/curate research data deposited into DataverseNO. Research data deposited in the top-level collection of DataverseNO are reviewed/curated by Research Data Service staff at UiT The Arctic University of Norway. If necessary, Research Data Service staff at UiT The Arctic University of Norway also give advice to Research Data Service staff at other DataverseNO partner institutions. The level of expertise of Research Data Service staff at partner institutions is not regulated by DataverseNO partner agreements. However, DataverseNO partner agreements require DataverseNO partners to fulfill all DataverseNO policies and guidelines, including the DataverseNO Curator Guidelines (see below). Research data deposited into special collections within the DataverseNO repository are reviewed/curated by Research Data Service staff who are highly proficient within the subject or discipline at stake. In TROLLing, review/curation is carried out by Senior Research Librarians responsible for language and linguistics at the University Library at UiT The Arctic University of Norway. If necessary, a scientific advisory board may be established for special collections within the DataverseNO repository; cf. TROLLing [1].

During review/curation, DataverseNO does not attempt to judge the scholarly quality of deposited datasets. As described in the DataverseNO Deposit Agreement [2], determination of the research quality is at the discretion of, and the responsibility of, the Long-Term Contact Person, as named in the metadata about the deposited dataset at stake.

Research Data Service staff review deposited datasets for alignment with criteria [3] [4] for depositing and/or to extend the metadata as needed to facilitate greater accuracy and discoverability. Both metadata and data files of deposited datasets are curated according to best practice. There are four areas to be checked: the uploaded files (both data and documentation), the registered metadata, the chosen license, and versioning, according to the checklist in the DataverseNO curator guidelines. Lack of compliance with the DataverseNO Deposit Agreement is communicated to the depositor and the dataset is returned for amendment. After finishing this review/curation process, the curator publishes the dataset.

Any changes in a dataset after its initial publication results in a new version of the dataset. Older published versions always remain openly accessible in DataverseNO. Published data can thus not be unpublished – with the only exception

being cases where access to the file(s) in a dataset or the entire dataset has to be removed. This process is regulated in the DataverseNO Preservation Policy [3].

During the review/curation process outlined above, the curator gives advice to the depositor about how to prepare and describe the dataset in order to obtain maximum re-usability of the data, as described in the DataverseNO Curator Guidelines [4]. The review/curation process may imply curation at all levels (A–D), including D-level with advice on formats for dates and numbers, or column headings. This review/curation process is carried out before the initial publication of datasets, and before any publication of a new version of a published dataset. For more information on the curatorial review process, please see the DataverseNO Curation Guidelines and the DataverseNO Accession Policy [5]. Datasets that are not compliant with the DataverseNO policies and guidelines are not published. If a curator identifies fundamental nonconformity with the DataverseNO policies and guidelines, and the depositor does not agree to make necessary changes, the curator addresses the problem by raising the issue within the curator community of DataverseNO to reach a conclusion. The conclusion is communicated to the depositor. If the reached conclusion is not accepted by the depositor, the issue is raised to the Board of DataverseNO. If applicable, the Board of DataverseNO may discuss the issue further with an advisory committee, before a final decision is made.

References:

[1] https://site.uit.no/trolling/people/

[2] DataverseNO Deposit Agreement (Data Deposit):

https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/

[3] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[4] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/

[5] DataverseNO Accession Policy (Quality Control):

https://site.uit.no/dataverseno/about/policy-framework/accession-policy/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

UiT The Arctic University of Norway (owner of DataverseNO) has an agreement [1] with BIBSYS as the national (Norwegian) DataCite DOI allocator agency, in order to make use of the DOI service from DataCite to assign persistent identifier to datasets and files in DataverseNO.

References:
[1] Agreement of cooperation for the provision of Digital Object Identifiers (DOI) and sharing research – confidential

document submitted with file name "Agreement_BIBSYS_DataCite.pdf" as part of the application submission.

Other Relevant Information.

The technical infrastructure of the DataverseNO repository is based on the Dataverse application [1]. The Dataverse application has its origin and base at Harvard University, and is currently used in about 50 installations worldwide. DataverseNO [2] as a national repository for research data is inspired by DataverseNL [3] in the Netherlands. Despite the parallel in naming however, DataverseNO is – unlike DataverseNL – not a network of individual repositories, but one repository with common policies and guidelines for operation and data stewardship.

DataverseNO makes use of the DOI service from DataCite to assign a persistent identifier to each dataset and to each file contained in a dataset. By this, DataverseNO contributes to the DataCite infrastructure with its metadata, and achieves increased visibility of its published datasets through the DataCite search and disseminating service [4]. The metadata in DataverseNO are open for harvesting from discovery services like Bielefeld Academic Search Engine (BASE) [5] and the Ex Libris Primo Central Index [6], and they are part of the global open access network enabled by the harvesting protocol OAI-PMH [7].

References:

[1] About Dataverse: https://dataverse.org/

[2] About DataverseNO: https://site.uit.no/dataverseno/about/

[3] DataverseNL: https://dataverse.nl/

[4] DataCite search and disseminating service: https://www.datacite.org/search.html

[5] Bielefeld Academic Search Engine (BASE): https://www.base-search.net/

[6] Ex Libris Primo Central Index: http://www.exlibrisgroup.com/products/primo-library-discovery/content-index/

[7] OAI-PMH: https://www.openarchives.org/pmh/

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

DataverseNO is a national, generic repository for open research data in the national research infrastructure of Norway. The service is owned by UiT The Arctic University of Norway and is operated and maintained following best practices for a sustainable data repository. The mandate and the specifications of this service are given by the University Management. The Board for DataverseNO is responsible for the repository [1], and ensures that the repository takes into account the interests and feedback of partner institutions, and the Designated Community. For more details about how responsibilities and roles are regulated within DataverseNO, see R0 and R5.

DataverseNO provides services for research data management according to best practice principles for secure archiving, preservation, and sustained, reliable and open access to research data in accordance with national and international guidelines and the FAIR principles for research data management.

An important part of the mission of DataverseNO is to acquire and preserve research data and provide access to them. DataverseNO is intended to provide maximum public access to unrestricted research data for the advancement of scholarship and the public good in ways that are consistent with the FAIR Data Principles [2] [3]. DataverseNO uses good archival practices to retain research data deposited into DataverseNO.

By the DataCite DOI minting requirements, UiT The Arctic University of Norway (owner of DataverseNO) is committed to secure archiving and data retrieval for at least 10 years after assigned DOI. Independent of this type of external

requirements, the intent of DataverseNO is to ensure access to archived data in a long-term perspective. UiT The Arctic University of Norway (owner of DataverseNO) has the responsibility to communicate to its partner institutions as well as individual depositors, the common guidelines for archiving and managing research data, and to keep these up to date in accordance with principles of best practice.

UiT The Arctic University of Norway (owner of DataverseNO) assume the obligation to ensure sound and reliable management of the repository service in accordance with the DataverseNO Preservation Policy [4].

References:

[1] Steering document for DataverseNO: https://site.uit.no/dataverseno/about/steering-documents/

[2] DataverseNO Policy Framework and Definitions: https://site.uit.no/dataverseno/about/policy-framework/

[3] FAIR Data Principles: https://www.force11.org/group/fairgroup/fairprinciples

[4] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

DataverseNO has agreements in place in order to regulate data deposit as well as data access and use.

Data deposit is regulated in the DataverseNO Deposit Agreement [1]. In order to deposit data into DataverseNO, the accept of this agreement has to be confirmed by the depositor on initial log-in to DataverseNO / signing up for a DataverseNO user account. The agreement covers the entire DataverseNO Policy Framework including several mutual rights and obligations that the depositor and UiT The Arctic University of Norway (owner of DataverseNO) accept regarding datasets to be deposited. The most important points are:
- The depositor holds the rights to grant the rights contained in the DataverseNO Deposit Agreement.
- If copyright terms for, or ownership of, the deposited data change, it is the responsibility of the Depositor to notify DataverseNO of these changes.
- By depositing data into DataverseNO, the depositor grants to UiT The Arctic University of Norway (owner of DataverseNO) the non-exclusive right to reproduce, translate, and distribute the Dataset in any format or medium worldwide and royalty-free, including, but not limited to, publication over the Internet.
- DataverseNO commits to preserving published Dataset in accordance with the DataverseNO Preservation Policy [2].

Data access and use are regulated in the DataverseNO Access and Use Policy [3]. Following the FAIR data principles, data in DataverseNO are released with a clear and accessible data usage license. As described in the DataverseNO Deposit Guidelines [4], the depositors are required to define a license for their dataset(s) at the time of deposit, and licensing information is displayed in the metadata for each dataset. The default Terms of Use for research data to be published in DataverseNO are Creative Commons CC0 – "No Right Reserved", accompanied by the following wording: "Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the archive". The CC0 license is considered best practice for optimal reuse of research data. However, individual collections within DataverseNO may choose to use a different default license with different Terms of Use. The individual depositors may in any case deviate from the default license by specifying different Terms of Use for their deposited dataset(s). When trying to download a dataset with another license than CC0 (preferably CC BY), the user is presented the actual license and terms, and must accept the conditions before downloading. Note that the default license CC0 in DataverseNO for reuse of data implies that there are no restrictions on reuse of the data. However, as is also stated in the Terms of Use, good scientific practice entails that proper credit is given via citation. In case the CC0 license is not suitable for a dataset, the depositor of the dataset is asked to contact Research Data Service staff at the DataverseNO partner institution responsible for the collection at stake for advice on which alternative license to choose. In line with the intention of DataverseNO to provide maximum public access to unrestricted research data, DataverseNO promotes licenses that are recommended for the re-use of research data, and only accepts licenses providing access to deposited data in one form or another. Guidance on choosing license is part of the curation process previous to the publication of datasets. In the case non-compliance with any access and use license other than CC0 (or equivalent) is discovered, DataverseNO confers with the contact person for the dataset. The use of the dataset must be terminated immediately at the initial demand by DataverseNO. If the use is not terminated, DataverseNO may bring action against the user.

References:
[1] DataverseNO Deposit Agreement: https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/

[2] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[3] DataverseNO Access and Use Policy, section Copyright and Licensing:

https://site.uit.no/dataverseno/about/policy-framework/access-and-use-policy/

[4] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data deposited into DataverseNO are managed according to best practice principles including secure archiving, preservation and continuous, reliable and open access to research data in accordance with national guidelines and EU principles for managing research data. The responsibility for compliance with these principles and guidelines is shared between UiT The Arctic University of Norway (owner of DataverseNO) and the individual partner institutions, and is regulated in the agreement on the use of DataverseNO (in Norwegian only). The key points in this agreement are listed below in English translation:

The partner institution is responsible for:

- the registration of research data in DataverseNO in compliance with current guidelines (DataverseNO Deposit Guidelines).

- the quality of metadata and the deposited research data from their own institution.

- their archived research data having a content that can be made openly available.

- clarifying the ownership and rights to the research data before archiving and publishing.

- user training and user support for employees and students at their institution.

UiT The Arctic University of Norway (owner of DataverseNO) is responsible for:

- the operation and management of DataverseNO.

- the adaption of common guidelines for DataverseNO (User Guides) to be in line with external requirements (for example from DataCite), best practice principles, and the functionality of the system.

- the integration of DataverseNO with the DOI service from DataCite so each archived dataset can be identified via DOI.

- secure archiving and access to the research data for a minimum of 10 years after assigned DOI, in accordance with the requirements from DataCite.

- preparing an institutional collection in DataverseNO for the partner institution.

- the allocation of data storage for the partner institution.

- the training and support of super users at the partner institution.

Responsibility

All depositors must accept the DataverseNO Deposit Agreement [1] prior to the archiving of data. This document provides to DataverseNO the non-exclusive right to reproduce, translate, and distribute the deposited items in any format or medium worldwide and royalty-free, including, but not limited to, publication over the Internet.

According to the Steering document for DataverseNO [2], UiT The Arctic University of Norway (owner of DataverseNO) is responsible for

- secure archiving and data retrieval for at least 10 years after the assignment of DOI.

- making the DataverseNO policies and guidelines known to administrators of DataverseNO.

- keeping the DataverseNO policies and guidelines up to date in accordance with best practice principles.

Partner institutions are responsible for ensuring that the DataverseNO policies and guidelines are applied to the institutional collections and the thematic sub-collections contained in these.

Continuity

UiT The Arctic University of Norway (owner of DataverseNO) is part of the national, governmental higher education and research system in Norway, as one of ten general state-founded universities under the ultimate responsibility of the Norwegian Ministry of Education and Research.

UiT The Arctic University of Norway has a long-standing record as a pioneer in promoting Open Access, Open Data and Open Science and has as a goal in its present strategy (2018-2022) to be national leading on Open Science. Thus, there is a strong commitment at the institution to support, prioritize and fund activities and services like DataverseNO, for the benefit of the institution.

The daily operations and the development of DataverseNO are managed by permanent staff from the University Library, the IT department and the Research administration at UiT The Arctic University of Norway, as part of their ordinary tasks within their organization, and based on defined responsibilities and roles agreed upon by the directors for the three organizational units, and approved by the university director. As such, DataverseNO is not a project or a separate organization or a corporate body, but a repository owned and operated by UiT The Arctic University of Norway, and offered as a service to other institutions.

As stated in the Steering document for DataverseNO, UiT The Arctic University of Norway (as owner of DataverseNO) commits to ensure the proper management and operation of the repository service in accordance with the responsibilities described in the document mentioned. The funding of DataverseNO consists of membership fees from partner institutions and internal funds and resources from UiT The Arctic University of DataverseNO (as owner of DataverseNO). The membership fee is based on established practices for common institutional services in the higher education sector in Norway, and includes fixed overhead expenses and volume pricing of storage services. Therefore, it is highly unlikely that UiT The Arctic University of Norway will close down DataverseNO. But if this unlikely scenario should take place, UiT The Arctic University of Norway (owner of DataverseNO) commits according to the DataverseNO Preservation Policy [3] to ensure that archived data is retained and transferred to approved repository/-ies in accordance with the agreement with DataCite for assignment of DOI to datasets in DataverseNO, before the service is discontinued. This will also be the preferred action for deposited data in an enduring perspective, as stated in the Steering document for DataverseNO. Datasets in the institutional collections are transferred to (a) certified general research data repository/-ies. Datasets in special collections are transferred to certified subject-relevant repositories after consulting the involved Designated Communities.

In addition and according to Norwegian legislation, research data from governmental sector will be transferred to the National Archives of Norway [4], securing long-term availability and accessibility of the data, in the case of closure of DataverseNO.

References:
[1] DataverseNO Deposit Agreement : https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/
[2] Steering document for DataverseNO : https://site.uit.no/dataverseno/about/steering-documents/
[3] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/
[4] The National Archives – https://www.arkivverket.no/en/about-us/the-national-archives-of-norway

Reviewer Entry
Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

DataverseNO is a repository for open research data – meaning that datasets must only contain unrestricted content with no private, confidential, or other legally protected information. DataverseNO may only make available content that is publicly distributable. This is part of the DataverseNO Deposit Agreement [1] that the depositor has agreed to before deposit, and the DataverseNO Accession Policy [2].

The depositor is solely responsible for the content deposited in DataverseNO, and shall not provide DataverseNO with any confidential or proprietary information that is required to be kept secret. By submitting content for deposing in DataverseNO, the depositor represents and warrants this to be in agreement with the General guidelines for research ethics, as well as subject-specific guidelines, from the Norwegian National Committees for Research Ethics [3] [4] [5]. DataverseNO may remove any content at any time if it does not comply with the DataverseNO Deposit Agreement.

Although the depositor is solely responsible for the content, Research Data Service staff will check and review deposited datasets before publishing (see requirements R0 Level of Curation Performed) [6]. This includes checking for compliance

with legal and ethical requirements, as well as with more general requirements in the DataverseNO Deposit Agreement. Any doubt or question concerning the compliance with the requirements mentioned above will be discussed with the depositor to secure compliance with the DataverseNO policies and guidelines before a dataset can be published. The Research Data Service staff taking care of the data curation are trained in performing the task by highly competent staff from the library at UiT The Arctic University of Norway, who also provide training and give courses [7] in various aspects of research data management, including management of research data with personal / sensitive information.

All employees of UiT The Arctic University of Norway, and the DatavereNO partner institutions are covered by the Norwegian Public Administration Act, section 13 and have signed a confidentiality agreement [8], ensuring that no confidential or personal information from their work (including DataverseNO) is disclosed.

DataverseNO requires that depositors define a license (see R2) for their dataset at the time of deposit, and licensing information is displayed in the metadata for each dataset. When trying to download a dataset with any other license than the default CC0, the user will be presented the actual license and terms (preferably CC BY), and must accept the conditions before downloading. In the case of non-compliance with any access and use license other than CC0 (or equivalent), the use of the dataset must be terminated immediately at the initial demand by DataverseNO. If the use is not terminated, DataverseNO may bring action against the user (see R2).

References:
[1] DataverseNO Deposit Agreement: https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/
[2] DataverseNO Accession Policy: https://site.uit.no/dataverseno/about/policy-framework/accession-policy/
[3] General guidelines for research ethics: https://www.etikkom.no/forskningsetiske-retningslinjer/Generelle-forskningsetiske-retningslinjer/general-guidelines-for-research-ethics/
[4] Norwegian National Committees for Research Ethics – Guidelines for Research Ethics in the Social Sciences, Law and the Humanities: https://www.etikkom.no/globalassets/documents/english-publications/guidelines-for-research-ethics-in-the-social-sciences-law-and-the-humanities-2006.pdf
[5] Norwegian National Committees for Research Ethics – Guidelines for Research Ethics in Science and Technology: https://www.etikkom.no/globalassets/documents/english-publications/guidelines-for-research-ethics-in-science-and-technology-2008.pdf
[6] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/
[7] Research data management training @ UiT: https://site.uit.no/rdmtraining/course-info/
[8] Public Administration Act, section 13: https://lovdata.no/dokument/NLE/lov/1967-02-10

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

I Organization

The organization of DataverseNO is described in the section Organization of DataverseNO [1] of the About page on the DataverseNO info site, and is discussed in detail below.

ORGANIZATIONAL DOCUMENTS
The organization, including repository structure, governance, data curation, and Designated Community, of DataverseNO is regulated in the following documents: Establishment of a Board for DataverseNO [2]; Mandate Board for DataverseNO [2]; Steering Document for DataverseNO [2]; DataverseNO Partner Agreements (attached to this application) including a data processor agreement; DataverseNO Policy Framework [3]; DataverseNO Administrator Guidelines [4]; DataverseNO Curator Guidelines [5]; DataverseNO Deposit Guidelines [6].

REPOSITORY STRUCTURE AND CONTENT
The repository structure of DataverseNO is discussed in R0. Below follows a brief overview of the collections in DataverseNO as of February 2020. For an updated overview, see the Support page [7] on the DataverseNO info site.

HVL Open Research Data
Institutional collection for Western Norway University of Applied Sciences. Collection launched in April 2019. No deposited

datasets. Two collection managers.

INN Open Research Data

Institutional collection for Inland Norway University of Applied Sciences. Collection launched in May 2019. Two published and three unpublished datasets. Two collection managers.

NMBU Open Research Data

Institutional collection for Norwegian University of Life Sciences (NMBU). Collection launched in October 2018. Eight published and nine unpublished datasets from researchers working within the following subjects: Agricultural Sciences; Business and Management; Medicine, Health and Life Sciences. Two collection managers.

NORD Open Research Data

Institutional collection for Nord University. Collection launched in June 2019. No deposited datasets. Three collection managers.

NTNU Open Research Data

Institutional collection for NTNU - Norwegian University of Science and Technology. Collection launched in January 2019. Seven published and four unpublished datasets from researchers working within the following subjects: Earth and Environmental Sciences; Medicine, Health and Life Sciences; Physics. Four collection managers

UiA Open Research Data

Institutional collection for University of Agder. Collection launched in August 2017. Six published and two unpublished datasets from researchers working within the following subjects: Computer and Information Science; Engineering; Medicine, Health and Life Sciences; Social Sciences. Four collection managers

UiB Open Research Data

Institutional collection for University of Bergen. Collection launched in June 2019. Three published and seven unpublished datasets from researchers working within the following subjects: Arts and Humanities; Medicine, Health and Life Sciences; Physics; Computer and Information Science; Earth and Environmental Sciences; Social Sciences. Three collection managers.

UiS Open Research data

Institutional collection for University of Stavanger. Collection launched in January 2020. No deposited datasets. Two collection managers.

UiT Open Research Data

Institutional collection for UiT The Arctic University of Norway. Collection launched in September 2016. 578 published and 29 unpublished datasets from researchers working within the following subjects: Agricultural Sciences; Arts and Humanities; Astronomy and Astrophysics; Business and Management; Chemistry; Computer and Information Science; Earth and Environmental Sciences; Engineering; Mathematical Sciences; Medicine, Health and Life Sciences; Physics;

Social Sciences. Eight collection managers.

TROLLing (The Tromsø Repository of Language and Linguistics)
Special collection for linguistic data and statistical code from linguists worldwide [8]. Collection launched in June 2014. 80 published and 20 unpublished datasets. Two collection managers.

GOVERNANCE

Ownership/hosting institution
DataverseNO is a repository owned and operated by UiT The Arctic University of Norway, and offered as a service to other institutions, and to individual researchers from research institutions in Norway. UiT The Arctic University of Norway is part of the national, governmental higher education and research system, as one of currently ten state-owned universities under the ultimate responsibility of the Norwegian Ministry of Education and Research (see also section II Funding below).

Board
The Board for DataverseNO have the overall responsibility for DataverseNO, with a mandate provided by the University Management of UiT The Arctic University of Norway [2].

Advisory Committees
Collections within DataverseNO may have their own advisory committees which give advice to the collection managers as well as to the Board of DataverseNO on high-level aspects of the operation and development of the collection at stake as well as the entire repository. Members of the Designated Community may raise any issues with representatives from the advisory committee of the collection at stake by contacting them directly. Currently, only TROLLing, a special collection in DataverseNO, has formally established an advisory committee, the TROLLing Scientific Advisory Board [8]. The TROLLing Scientific Advisory Board provide their advice to the managers of TROLLing.

The operation of institutional collections is part of the research support services and the institutional management at the DataverseNO partner institutions. Partner institutions have well-established venues in place where research support units, such as the University Library, discuss issues with representatives from the different research communities at the institution. Feedback from such discussions is provided to the managers of the institutional collections. On their part, managers of institutional collections discuss advice and feedback from the user groups of their institutional collections in the Advisory Committee for DataverseNO. This committee, illustrated with the blue box in the middle of the GOVERNANCE section of the DataverseNO Organization Chart, consists of representatives from all DataverseNO partner institutions (usually the collection managers), and the managers of DataverseNO. The members of the DataverseNO Advisory Committee meet at least twice a year to discuss issues concerning the organization of DataverseNO, including governance, policies and guidelines, repository structure and operation (including functionality), data curation, and issues raised by the Designated Community. Requests and advice from the DataverseNO Advisory Committee are communicated to the Board of DataverseNO and to the managers of the institutional collections by the

DataverseNO Repository Management.

Management
Repository Management
The daily management and operation of DataverseNO are carried out by permanent staff from the Library, the IT department and the Research administration at UiT The Arctic University of Norway, as part of their ordinary tasks within their organization, based on defined responsibilities and roles agreed upon by the directors for the three organizational units, and approved by the university director.

The repository management of DataverseNO consists of three permanent staff members from the UiT Library. They are responsible for the management, maintenance, development and the daily operation of the repository, and they take care of the DataverseNO policies and guidelines, communication with the Board of DataverseNO, communication with and training of collection managers, the operation of the DataverseNO Advisory Committee, the configuration of the repository, establishment and configuration of institutional collections, user management, the implementation of new functionality and procedures to be used in the repository, preservation planning, and the certification of the repository. In addition, the DataverseNO repository management is responsible for the management of the top-level collection of the repository.

The technical operation and maintenance of the repository is carried out by two computer engineers from the UiT Library, and two computer engineers from the UiT IT department. The computer engineers at the library are responsible for the installation, customizing, and upgrading of the repository application. Note that the Dataverse application is only slightly customized for use in DataverseNO. The computer engineers at the IT department take care of the secure and sustainable operation, back-up, and upgrading of the servers used to run the repository application as well as of the customization and infrastructure used for federated authentication.

In addition, the management of DataverseNO involves one staff member from the Research administration at UiT, and one staff member from the IT department at UiT. They both work together with the service management. The staff member from the Research administration is responsible for the alignment of DataverseNO with the UiT policies and strategic framework, whereas the staff member from the IT department is responsible for the strategic development of IT infrastructure relevant to DataverseNO.

Collection Management
Institutional Collections
The managers of institutional collections within DataverseNO are responsible for the management and operation of the collection, including compliance of the institutional collection and underlying sub-collections with the DataverseNO policies and guidelines, user management of collection curators, training of and communication with collection curators, establishment and configuration of sub-collections, communication with DataverseNO repository management, communication with the management at the partner institution, representing the institutional collection in the DataverseNO Advisory Committee. The management of the institutional collections in DataverseNO are all Research Data Service staff members at the partner institutions. Each collection has at least two managers.

Special Collections

The managers of special collections have many of the same responsibilities as institutional collection managers, but limited to the thematic scope of the collection. They communicate with the advisory committee for the collection – if applicable. Currently, TROLLing is the only special collection in DataverseNO. TROLLing has two managers.

DATA CURATION

Curation of data deposited in institutional collections is the responsibility of the partner institutions, and is carried out by Research Data Service staff at these institutions. Datasets deposited in the top-level collection are curated by Research Data Service staff at UiT The Arctic University of Norway (owner of DataverseNO). Datasets deposited in special collections are curated by Research Data Service staff specialized in the subject(s) covered by the collection. The currently only special collection, TROLLing, is curated by subject librarians for linguistics at UiT The Arctic University of Norway. All data curation in DatavereNO is carried out by staff members employed at the partner institutions. Typically, these data curators are mainly permanent staff working as subject librarians or as research support advisers at the library or in the different faculties and/or institutes at the DataverseNO partner institutions. Many of the data curators have PhD level education within the research disciplines for which they are providing support services. In addition, they have been, and are continuously, trained in research data management, and they have in-depth knowledge of data stewardship within the research disciplines they are set to support. Furthermore, they keep themselves up to date with developments of both general and subject-specific standards and best practices for research data management. The combination of being trained as both researchers and research data management specialists makes them highly qualified for supporting the data stewardship of their institutional collection within DataverseNO. The responsibility for the management and curation of the institutional collections is placed at the different partner institutions precisely because they know the needs and therefore are best suited to serve the research communities at their institution, and thus, the user groups of the respective institutional collections within DataverseNO.

Data curators are responsible for ensuring that data published in each collection within DataverseNO (including the top-level collection) are curated according to the DataverseNO policies and guidelines, and in line with best practice recommendations and the needs of the different user communities at stake (see R0 on partner agreements and Designated Community). Curators communicate with the different user communities represented in the collection(s) they curate, e.g. during curation, but also through other channels and in other venues. Curators also communicate with the management of their collection, and with curators of other collections within DataverseNO building the DataverseNO Network of Expertise. This network of curators covers the different aspects of data curation, including metadata, file formats, and licensing. In addition to enabling knowledge and experience exchange, this network also makes sure that curation practices across the repository are aligned with the DataverseNO policies and guidelines, and also seeks to align curation practices across institutional collections from different partner institutions containing data from the same or similar scholarly disciplines. At collection launch, each partner institution starts off with at least two curators. UiT has currently eight curators. See also discussion about resource scaling in section II Funding below.

DESIGNATED COMMUNITY

See R0.

II Funding

DataverseNO is organized in a way that ensures sufficient funding for the operation and further development of the repository in a long-term perspective.

To be noted on a general level, both the owner institution and the partner institutions are state-owned universities and thus part of the national, governmental higher education and research system and under the ultimate responsibility of the Norwegian Ministry of Education and Research [9]. They are all reputable institutions that have existed for many decades – though in some cases not under their current name. Thus, they all are organized and funded in a way that ensures the operation of sustainable services for higher education and research in an enduring perspective. Also, all institutions involved in DataverseNO have recognized Open Science as an important issue in their missions.

Still on a general level, it is also of utmost importance to make clear that – as is the case for any other sustainable service – both the owner institution and the partner institutions of DataverseNO allocate their funding and resources to the operation and development of DataverseNO on a scalable basis, but always to a sufficient extent in order to completely fulfill their commitments at any time. This means, e.g., that a partner institution does not allocate all their research support staff to the operation of their institutional collection within DataverseNO right from the establishment of the collection. Allocation of resources on a scalable basis means that necessary funding and staff are allocated gradually as data deposit into the collection increases. This scalable model has proved to be very successful and sustainable in the development and operation of similar services at higher education and research institutions in Norway.

Furthermore, although the resources needed e.g. for data curation increase as more researchers at DataverseNO partner institutions choose to deposit their data into DataverseNO, we expect, and have already experienced, that the average time used on data curation per dataset will decrease as researchers become more proficient in research data management the more datasets they have deposited into the repository and the more research data management training they have received at the partner institution or elsewhere. The details presented below should be understood on this general background.

Owner of DataverseNO
UiT The Arctic University (owner of DataverseNO) has a long-standing record as a pioneer in promoting Open Access, Open Data and Open Science in Norway, and has as a goal in its present strategy (2018-2022) to be nationally leading in Open Science [108]. Thus, there is a strong commitment at the institution to long-term support, strategic priority and sustainable funding of activities and services like DataverseNO, for the benefit of the institution. In particular, as described

in the official and publicly available Steering Document for DataverseNO [2], UiT commits to the partner institutions and the Designated Community of DataverseNO to ensure the proper management and operation of DataverseNO in a long-term perspective, and in accordance with the responsibilities described in this document.

Partner institutions

By signing the partner agreement, the partner institutions of DataverseNO commit to operate their institutional collections according to DataverseNO policies and guidelines. Although not explicitly mentioned in the agreement this implies that they have to ensure sufficient funding and resources as well as sufficiently qualified staff to fulfill these requirements at any time.

Funding model of DataverseNO

Before DataverseNO was established as a national generic repository for open research data, the repository served as a generic institutional repository for UiT The Arctic University of Norway, operated and funded by the institution. As the founder and owner of DataverseNO and due to the institutional need for such a service, UiT The Arctic University of Norway take the responsibility for the basic funding of the repository. The partner membership fees cover UiTs overhead expenses for offering DataverseNO to their partner institutions. These overhead expenses are related to the management, the operation, and the development of the repository, but not to data curation of any sort – since data curation is the responsibility of the partner institutions. Each partner institution covers their expenses for necessary staff resources, competence building and attending meetings, etc.

Allocation of staff

Currently (as of February 2020), the following staff resources are allocated to the operation and further development of DataverseNO. See discussion of scalable model above.

Repository Management and Operation

- Three permanent staff members from the UiT Library responsible for service management; approx. 2 FTEs
- Four permanent staff members from the UiT Library and the UiT IT department for technical operation and maintenance of the repository; approx. 0.75 FTEs
- One permanent staff member from the Research administration at UiT for alignment with UiT strategy; approx. 0.1 FTEs
- One permanent staff member from the IT department at UiT for strategic development of IT infrastructure; approx. 0.1 FTEs

Collection Management and Operation

Institutional Collections

- Two permanent staff members for collection management at each of the following seven partner institutions: Inland Norway University of Applied Sciences, Nord University, Norwegian University of Life Sciences, University of Agder,

University of Stavanger, University of Bergen, Western Norway University of Applied Sciences
- Three permanent staff members for collection management at each of the following two partner institutions: NTNU - Norwegian University of Science and Technology, UiT The Arctic University of Norway

As is apparent from the overview (see section I Organization above) of deposited and published datasets in the different institutional collections, apart from UiT, all collections are still in their establishing phase. The number of FTEs for the management of these collections are currently approx. 0.2 for each collection. At UiT the corresponding number is approx. 0.5 FTEs.

Special Collections
TROLLing has two permanent staff members for collection management. The current allocation of FTEs for the management of TROLLing is approx. 0.3 FTEs.

Data Curation
- At least two permanent staff members for data curation of each institutional collection of DataverseNO. Institutions with more staff members: Nord (3), UiB (3), NTNU (4), UiA (4).
- Eight permanent staff members for data curation of the institutional collection for UiT and the DataverseNO top-level collection
- Two permanent staff members for data curation of TROLLing

The current allocation of FTEs for the management of these collections are as follows:
- HVL Open Research Data: approx. 0.1 FTEs
- INN Open Research Data: approx. 0.1 FTEs
- NMBU Open Research Data: approx. 0.3 FTEs
- NORD Open Research Data: approx. 0.1 FTEs
- NTNU Open Research Data: approx. 0.4 FTEs
- UiA Open Research Data: approx. 0.3 FTEs
- UiB Open Research Data: approx. 0.3 FTEs
- UiS Open Research Data: approx.. 0.1 FTEs
- UiT Open Research Data (incl. special collection (TROLLing), and top-level collection): approx. 4 FTEs

To summarize, the staff resource allocated to the operation and development of DataverseNO amount to a total of approx. 52 permanent staff members accounting for approx. 11 FTEs. (Note that in quite a few cases a single staff member may have different roles in DataverseNO; the total number of staff members reported above applies to unique staff members.)

In addition to the commitment described above, each partner institution (incl. UiT) allocates an increasing amount of staff

resources to provide research data management training for research support staff, researchers, and students at the institution (see section III below). Although these resources are not included in the numbers above, they undoubtedly benefit the operation of DataverseNO as they – among other things – contribute to increase the pool of qualified Research Data Service staff who may be allocated to the operation and curation of institutional and special collections within DataverseNO.

III Training and Professional Development

All management, operation, data curation, and development of DataverseNO are carried out by permanent research support staff members at the DataverseNO owner institution and the DataverseNO partner institutions. As explained above, these institutions are all higher education and research institutions in Norway. As such, they place great emphasis on ongoing training and professional development of their employees as part of their ordinary work tasks. The Research Data Service staff involved in the operation of DataverseNO keep themselves up to date on developments within the scholarly disciplines they provide research support services for, as well as on standards and best practice recommendations for research data stewardship. They regularly attend workshops, webinars, training courses, conferences, and other training events on research data stewardship, both in Norway and abroad. Among the conferences and training events attended in the past five years are the annual meeting of the International Association of Social Science Information Services and Technology (IASSIST), the plenary meetings of the Research Data Alliance (RDA), Dataverse Community Meetings, the International Data Curation Conference (IDCC), in addition to several national and Nordic events, e.g. workshops on ethical issues in research data management organized by NSD - Norwegian Centre for Research Data. Recently, two members of the Research Data Staff at UiT The Arctic University of Norway received a GO FAIR Readiness Certificate after attending a 4-day course on FAIR data stewardship [11].

In addition to the training activities mentioned above, UiT The Arctic University of Norway – as the owner of DataverseNO – take special responsibility for keeping the repository and the involved personnel up to date with any matters relevant to the proper operation of the repository in accordance with international standards and best practice recommendations. UiT offers regularly training courses on how to manage research data, included training in how to archive and share research data. The courses are run by Research Data Service staff from the Library (main contributor), the IT department, and the Research administration at UiT. The courses cover all levels from basic research data management to different advanced topics [12]. The contents of these training activities as well as the competence and the experiences from these activities are shared with the partner institutions of DataverseNO. Also, UiT and/or its personnel are members and actively participate in international bodies including CLARIN, Liber, and RDA (see also R6).

Expertise in data curation and other aspects of data stewardship relevant to the proper operation of DataverseNO is shared through different channels within DataverseNO, e.g. via the DataverseNO Advisory Committee, and the Network of Expertise established between the data curators of the different collections within DataverseNO.

Furthermore, the newly established national RDA group for Norway will play an essential role in the training of Research

Data Service staff at higher education and research institutions in Norway, and also in the alignment of practices for research data management in Norway with international standards and best practice recommendations [13]. One of the repository managers of DataverseNO and one of the managers of the institutional collection of NTNU are involved as key personnel of the Norwegian RDA group, and we expect the activities and outcomes of the group to be of great benefit for DataverseNO.

IV Range and Depth of Expertise

The owner institution as well as the partner institutions of DataverseNO have as their main mission to provide high-quality services in the higher education and research sector, and they have done so for many decades. The research support staff at these institutions have high-level and in-depth expertise on the different scholarly subjects for which they offer services. The Research Data Service staff responsible for the operation and development of the DataverseNO repository and its collections consist of a range of individuals who are highly qualified for their tasks.

The repository and collection management is carried out by permanent library staff members with at least graduate education in addition to training in research data stewardship. They have long experience from developing and operating research support services, e.g. Open Access publishing services. The technical aspects of the repository are taken care of by computer engineers with graduate educations and long experience from developing, operating, and maintaining the technical infrastructure of research support services.

Data curation in DataverseNO is carried out by permanent library staff members at the owner institution and at the partner institutions. Most of these staff members are (Senior) Research Librarians, many of which with PhD (level) education within the scholarly subjects they curate research data from. In addition, they have been – and are continuously – trained in standards and best practices for research data management. The DataverseNO curators also share and align their expertise and practice through the DataverseNO Network of Expertise.

Although the mission of DataverseNO – with the possible exception of special collections – is to be a national GENERIC repository for open research data, the repository strives to provide subject-specific expertise as far as possible; see R6, R8, and R11. This is why, as a main rule, data deposited into institutional collections or into the top-level collection of DataverseNO are curated by Research Data Service staff who are subject specialists in addition to being trained in research data management. Special collections of DataverseNO are without exception managed and curated by permanent Research Data Service staff who are subject specialists.

References:
[1] Organization of DataverseNO: https://site.uit.no/dataverseno/about/#organization-of-dataverseno
[2] DataverseNO Steering Documents: https://site.uit.no/dataverseno/about/steering-documents/
[3] DataverseNO Policy Framework: https://site.uit.no/dataverseno/about/policy-framework/

[4] DataverseNO Administrator Guidelines:

https://site.uit.no/dataverseno/admin-en/administration-dataverseno/

https://site.uit.no/dataverseno/admin-en/administration-collections/

[5] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/

[6] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

[7] DataverseNO support page: https://site.uit.no/dataverseno/support/

[8] TROLLing Scientific Advisory Board: https://site.uit.no/trolling/people/

[9] State-owned universities and university colleges in Norway: https://www.regjeringen.no/en/dep/kd/organisation/kunnskapsdepartementets-etater-og-virksomheter/Subordinate-agencies-2/state-run-universities-and-university-co/id434505/

[10] Strategic plan for UiT The Arctic University of Norway 2014-2022:

https://en.uit.no/om/art?p_document_id=377752&dim=179033

[11] FAIR data stewardship course: https://indico.neic.no/event/56/

[12] Research data management training @ UiT: http://site.uit.no/rdmtraining/?lang=en

[13] RDA Group Norway: https://rd-alliance.org/groups/rda-norway

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

The implicit expectation in the Partner Agreement that Partner Organisations will 'ensure sufficient funding and resources as well as sufficiently qualified staff to fulfil these requirements at any time.' does not adequately address this Requirement. The Partner Organisations' funding and resource commitment should be stated explicitly in the future as part of the Partner Agreement.

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry
Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

DataverseNO has infrastructure and procedures in place to secure the continuous advice and feedback of experts in the fields relevant for the proper and sustainable operation of the repository in compliance with standards and best practice recommendation for research data management, both in general and, where applicable, within the different scholarly disciplines represented in DataverseNO.

In-house expertise
As discussed in R0 and R5, the DataverseNO repository and the data deposited into the repository are managed and curated by permanent Research Data Service staff at the DataverseNO owner institution and at the DataverseNO partner institutions. These institutions have the range and depth of expertise necessary to ensure compliance with the DataverseNO policies and guidelines, and to take care of the interests of the Designated Community of the repository.

Although the mission of DataverseNO – with the possible exception of special collections – is to be a national GENERIC repository for open research data the repository strives to provide subject-specific expertise as far as possible; see R6, R8, and R11. This is why, as a main rule, data deposited into institutional collections or into the top-level collection of DataverseNO are curated by Research Data Service staff who are subject specialists in addition to be trained in research data management. Special collections of DataverseNO are without exception managed and curated by permanent Research Data Service staff who are subject specialists.

In addition to the training activities provided by UiT The Arctic University as well as at the different partner institutions, the collections managers use the DataverseNO Advisory Committee, and the data curators use their network of expertise, to share, keep up to date, and align their expertise.

Through its network of participating and collaborating institutions, DataverseNO has access to a pool of experts in the field of research data management. Participating institutions in DataverseNO are all research institutions, hosting a wide range of experts including the full range of academic subjects represented in DataverseNO, IT experts from the IT departments, as well as legal experts from the institutions' administrative departments. All participating institutions in DataverseNO are organized in faculties and institutes which have their own boards that advise and decide on important and strategic matters relevant for the operation of these organizational units, as well as for the research communities that are part of these units. Since the institutional collections of DataverseNO are part of the organizational structure of the Dataverse partner institutions the feedback and expertise from these bodies are integrated directly into the DataverseNO-internal flow of communication and network of expertise, with the collection managers as the main liaison.

A special collection is established at a DataverseNO partner institution with research expertise within the field of study at

stake, on request from a user group. In the unlikely case that the field of study is closed down at the institution, the responsibility for the collection is transferred to another DataverseNO partner institution with the relevant expertise. If this is not possible, the data in the collection are transferred to another subject-relevant repository (preferably) or a generic repository. See also section Continuity in R3.

External expertise

DataverseNO and its partner institutions also have access to advice from external experts, both nationally and internationally.

UiT The Arctic University of Norway (owner of DataverseNO) as well as all other DataverseNO partner institutions are collaborating with the Norwegian Centre for Research Data (NSD), especially on the management of personal data [1]. Research Data Service staff managing collections and curating datasets in the DataverseNO repository confer with experts at NSD in case they need special advice on issues regarding personal data.

Several Research Data Service staff members at UiT The Arctic University of Norway (owner of DataverseNO) and other DataverseNO partner institutions participate in and contribute to several interest and working groups of the Research Data Alliance (RDA) [2]. Examples are the Libraries for Research Data IG (member), the Data Citation WG (member), the Linguistics Data IG (co-initiator), and the DMP Common Standards WG (member). Through the extensive RDA network, DataverseNO managers and curators have access to expertise covering essentially all topics within research data management. Several of the participating institutions in DataverseNO are also represented in the recently established national RDA group for Norway [3]. One of the main goals of this group is the coordination and collaboration in matters of Research Data Management in Norway.

As the owner of DataverseNO, UiT The Arctic University of Norway has access to expertise from DataCite [4] in questions concerning citation metadata compliance.

UiT The Arctic University of Norway (owner of DataverseNO) is participating and contributing to the global Dataverse User Community. The community has recently been formally organized in the Global Dataverse Community Consortium [5], which aims at providing a collaborative venue for institutions to leverage economies of scale in support of Dataverse repositories around the world.

Collections within DataverseNO may establish their own advisory boards. This has been done for TROLLing [6]. The members of the TROLLing Scientific Advisory Board contribute with top-level scientific expertise to the operation and development of TROLLing. TROLLing is also participating in CLARIN - European Research Infrastructure for Language Resources and Technology [7]. Being part of CLARIN, UiT The Arctic University of Norway (owner of DataverseNO) is together with a number of European key stakeholders participating in the Social Sciences & Humanities Open Cloud (SSHOC), the SSH part of the European Open Science Cloud (EOSC) [8]. Through the EOSC network, not only TROLLing, but the entire DataverseNO repository, has access to top-level expertise in providing research data management services to all of its Designated Community.

Although, as mentioned initially, DataverseNO – apart from special collections (currently TROLLing) – is a national GENERIC repository for open research data the repository strives to provide subject-specific expertise as far as possible. Therefore, access to top-level scientific expertise in DataverseNO is not restricted to the field of Linguistics as described in the previous section. In addition to the in-house scientific expertise described in the section "In-house expertise" above DataverseNO has through its participating institutions access to national expertise in all the scholarly subjects represented at all higher education and research institutions in Norway. All these institutions – including all participating institutions in DataverseNO – are members of Universities Norway/Universitets- og høgskolerådet (UHR), which is a cooperative body for 33 accredited universities and university colleges in Norway [9]. UHR works with research and higher education policy and coordination within the university and college sector, both at the national and international level.

Communication with experts for advice

Our preferred form of communication with the experts mentioned above is through direct contact, e.g. by email, at on-line or personal meetings or through on-line community fora. UiT The Arctic University (owner of DataverseNO) assists all DataverseNO partner institutions in keeping up to date on relevant changes and enhancements in the field of Research Data Management.

Communication with Designated Community

The infrastructure and procedures that DataverseNO uses to communicate with its Designated Community for feedback are discussed in the section "Brief Description of the Repository's Designated Community" in R0.

References:

[1] Norwegian Centre for Research Data (NSD): http://www.nsd.uib.no/nsd/english/index.html

[2] Research Data Alliance: https://www.rd-alliance.org/

[3] RDA Group Norway: https://rd-alliance.org/groups/rda-norway

[4] DataCite: https://www.datacite.org/

[5] The Global Dataverse Community Consortium: http://dataversecommunity.global/

[6] TROLLing Scientific Advisory Board: https://site.uit.no/trolling/people/

[7] CLARIN - European Research Infrastructure for Language Resources and Technology: https://www.clarin.eu

[8] Social Sciences & Humanities Open Cloud (SSHOC): https://www.sshopencloud.eu/

[9] Universities Norway/Universitets- og høgskolerådet (UHR): https://www.uhr.no/en/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

When digital objects are uploaded to DataverseNO, the system runs two integrity checks during file ingest. Universal Numerical Fingerprint (UNF) checksums [1] are applied as indicators to be used to verify that no changes have been made to tabular data in the dataset. MD5 checksums [2] are applied to each file as indicators to be used to verify that the files have not been altered. The storage systems are renewed every 6-8 years which minimizes the risk for long-term deterioration of storage media. The transfer of data from old to new storage systems includes checks for bit-correctness of all data. See also R9.

On deposit, DataverseNO Research Data staff check data files and metadata for completeness and integrity and require, as needed, changes in data files and/or metadata from depositors. As an important part of the documentation, a ReadMe file must accompany each dataset, with a description of how to (re)use the dataset, including a statement of the completeness, or the limitations, of the dataset; see the DataverseNO Deposit Guidelines [3]. This ReadMe file is reviewed by Research Data Service staff before the dataset is published.

Changes to data files and metadata of published datasets are logged in the Dataverse version control report. Any change creates a new version of the dataset, including documentation of what has been changed and by whom. Minor additions or revisions of the metadata yield a decimal version number change. Additions of new data or other major alterations of

existing data yield a change in the integer version number. Previous versions of datasets remain always openly accessible. Changes between subsequent versions of datasets are openly documented through version control. Any change to published datasets is subject to review by Research Data Service staff.

Data authenticity

Depositors may apply changes to their data published in DataverseNO as described above. The procedures for such data changes are communicated to depositors through the DataverseNO Deposit Guidelines [3], and to Curators through the DataverseNO Curator Guidelines [4]. In order to ensure the long-term preservation and usability of published datasets, changes to data may also be applied by Research Data Service staff; see R10. The rationale and procedures for such changes are regulated in and communicated to depositors through the DataverseNO Preservation Policy [5].

According to the DataverseNO Deposit Guidelines, depositors have to provide documentation about the creation of the data, and how the data can be used. This documentation must be provided in a ReadMe file that is deposited together with the data. In addition, provenance information may be entered into the metadata schemas provided by the repository software. Provenance information of the latter type is provided at file level and accepted in two forms: as a provenance file in JSON format and following W3C standards, and/or as a free-text provenance description.

Links to metadata records, and to other datasets that are related to the dataset in question, are maintained through the Related dataset metadata field. DataverseNO is following closely the project of integration between DataCite and CrossRef, that will enable automatic linking between related datasets and publications [6]. Keywords in the metadata record of deposited datasets may refer and link to metadata standards, e.g. controlled vocabularies. These links are reviewed by Research Data Service staff before the dataset is published. To ensure sustainability, the metadata elements from such external sources are in any case always stored in the metadata record of the dataset in question. The links are thus only meant for reference.

The version control system described above provides information about the essential properties of different versions of the same file.

The names of the depositor and the curator are registered automatically during data deposit and curation, and the identity of the depositor is verified by required log-in through the Norwegian national Lightweight Directory Access Protocol (LDAP) system (Feide) [7].

References:
[1] Universal Numerical Fingerprint (UNF): http://guides.dataverse.org/en/latest/developers/unf/index.html
[2] Checksum (MD5): https://en.wikipedia.org/wiki/MD5
[3] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/
[4] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/
[5] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/
[6] Scholix: http://www.scholix.org/
[7] Feide: https://www.feide.no/introducing-feide

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

DataverseNO is a Norwegian national, generic repository for open research data. The DataverseNO Accession Policy [1] explains what DataverseNO can accept for archiving. The DataverseNO Accession policy as well as the DataverseNO Deposit Guidelines [2] also include guidelines on how to select data for archiving.

Data accepted for archiving in DataverseNO are in digital formats, and they are either generated through the course of a research project and/or deposited with an expectation that public availability will allow the data to be used for research purposes. As a GENERIC repository, the collection development policy of DataverseNO does not put any limitations on the field of study represented in the data to be deposit. However, special collections within DataverseNO may in addition have requirements on the subject area of the research data to be deposited. Currently, TROLLing is the only special collection in DataverseNO. TROLLing only accepts research data from linguistics / language studies.

Although the mission of DataverseNO – with the possible exception of special collections – is to be a national generic repository for open research data the repository strives to provide subject-specific expertise as far as possible; see also

R6, and R11. This is why, as a main rule, data deposited into institutional collections or into the top-level collection of DataverseNO are curated by Research Data Service staff who are subject specialists in addition to be trained in research data management. Special collections of DataverseNO are without exception managed and curated by permanent Research Data Service staff who are subject specialists.

After deposit, each dataset is curated by Research Data Service staff before publication, to ensure compliance with the DataverseNO Accession Policy [1] [3], and the DataverseNO Deposit Guidelines [2], regarding completeness, organization and documentation of the data. If necessary, Research Data Service staff communicate with depositors to make the dataset compliant with these policies and guidelines.

Depositors must make a selection or appraisal of which files to be deposited in order for the dataset to be complete and understandable. As a general rule, enough data must be provided for others to be able to understand and replicate the study or otherwise (re)use the deposited data. Decisions on data selection and completeness should preferably be based on general discussions in the institutional, national and international research communities about what is appropriate and what is considered good practice within the discipline in question. This approach is fully in line with the recommendations in the National policy for research data management in Norway, which states that questions regarding what data researchers should make openly available "are questions that researchers themselves have to decide on through discussions in the institutional, national and international research communities about what is appropriate and what is considered good practice within different subject areas" [4].

The DataverseNO Accession Policy requires depositors to provide enough data and metadata (included a ReadMe file) so that others can understand and (re)use the data. Our Deposit Guidelines describe in more detail how datasets have to be prepared and documented according to best practice before they are deposited to the repository. Datasets submitted to the repository are curated by Research Data staff before they are published. The curation process assures as far as possible that the deposited datasets are complete and understandable. Datasets not complying with these requirements are returned to the author together with requests to adjust and/or better describe or document the dataset in order to comply with our guidelines. The curation procedures are described in the DataverseNO Curator Guidelines [5].

The DataverseNO Accession Policy requires deposited datasets to be in (a) preferred file format(s) to facilitate long-term preservation. The DataverseNO Deposit Guidelines include a list of preferred file formats for common document types. Adherence to preferred file formats is part of the curation process, as described in the DataverseNO Curator Guidelines. File formats not included in the list, will be assessed during the curation process. Research Data Service staff closely follow best practice in the field of preservation in digital archiving in order to be able to advise depositors on the sustainability of different data formats.

As a main rule, DataverseNO requires data to be deposited in their original file format in addition to a preferred file format (if the original is not in a preferred format), as described in the DataverseNO Deposit Guidelines. If data are deposited in non-preferred file formats only, the dataset is returned to the depositor together with a request to provide the data in preferred file formats as well. The DataverseNO Deposit Guidelines also give advice on how to convert data files from non-preferred file formats into preferred file formats. However, if the research data are represented in a non-preferred file

format that is commonly used by the research community at stake, and the file format cannot be converted into a preferred format, DataverseNO accepts the data for deposit with the limitations this implies for long-term preservation; see R10.

If – after the curation process – the depositor is not able to provide data that are sufficiently complete and sufficiently documented they cannot be published in DataverseNO. For data that have been accepted and published, the DataverseNO Deposit Agreement grants DataverseNO the right to amend the metadata as well as convert and migrate data files to any medium or format for the purposes of (long-term) preservation [6]. The measures for long-term preservation of datasets published in DataverseNO are described in R10. In case the metadata provided in a published dataset at a later stage nevertheless turn out to be insufficient for long-term preservation Research Data Service staff responsible for the curation of the dataset in question will attempt to obtain more information about the dataset from the depositor in order to update the preservation metadata about the dataset. If this information cannot be obtained from the depositor Research Data Service staff will ask for expertise help from the Designated Community and the experts described in R6.

References:

[1] DataverseNO Accession Policy: https://site.uit.no/dataverseno/about/policy-framework/accession-policy/

[2] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

[3] DataverseNO Policy Framework and Definitions: https://site.uit.no/dataverseno/about/policy-framework/ (see section "Quality Commitment")

[4] National policy for research data management in Norway (12/2017): https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf (p. 26, Norwegian only; English translation given in answer to R0 above)

[5] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/

[6] DataverseNO Deposit Agreement: https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The DataverseNO infrastructure is operated and managed by the IT department at UiT The Arctic University of Norway (owner of DataverseNO), and has the same level of service quality and operational security as all other application services at the institution provided by the IT department. The infrastructure and services are revised yearly according to the IT department quality control system. The quality control system is based on the following standards for quality management systems [1]:

NS-EN ISO 9000:2006 - Grunntrekk og terminologi (Basics and terminology)

NS-EN ISO 9001:2008 - Krav (Demands)

NS-EN ISO 9004:2009 - Kvalitetsstyring som metode (Managing for the sustained success of an organization -- A quality management approach)

NS-ISO 10005:2005 - Retningslinjer for kvalitetsplaner (Quality management systems -- Guidelines for quality plans)

NS-ISO/TR 10013:2001 - Retningslinjer for dokumentasjon av system for kvalitetsstyring (Guidelines for quality management system documentation)

NS-EN ISO 19011:2011 - Retningslinjer for revisjon av styringssystemer (Guidelines for auditing management systems)

All access to the management interfaces are restricted both through network segmentation, protocol encryption and authorization only for the personnel required for operating the infrastructure. All data centers have physical security implemented with key-cards and access restrictions limited to necessary staff.

UiT (owner of DataverseNO) is committed to sustaining an effective digital preservation infrastructure for its digital collections, which includes the adequate provision of appropriate technologies [2]. The DataverseNO Preservation Policy [3] describes the technological sustainable storage of all content in the repository. Datasets deposited into DataverseNO utilize the centralized back-end storage and management services at UiT. This is a common storage and management infrastructure for digital collections of enduring value to UiT, covering digitized and born-digital books, manuscripts, photographs, audio-visual materials, scholarly publications, and research data.

DataverseNO is running on UiT's centralized storage and virtualization infrastructure which also hosts the accounting and

payroll systems for the whole institution. Everything is backed up using an enterprise class backup system with retention policies ensuring that multiple copies are maintained of all data in the system. The underlying hardware is mirrored between two datacenters in separate buildings on the UiT campus.

The backup routine builds on a daily backup with a snapshot of the data and the metadata, as well as the whole VMWare-server. The backup consists of a full snapshot of the server each 90th day followed by a daily incremental snapshot with an integrity check, until the next full backup. In this way, the state of the virtual machine can be restored 90 days back in time, or files / databases can be retrieved 90 days back in time. The backup-data are stored in a separate datacenter (separate building) 500 m from where the production server runs.

Recovery time depends on the amount of data. Currently (850 GB), it will probably take up to 1 hour to take a full restore of the server, including the OS-system as well as the application DataverseNO with all the data. A file or partly restore will normally take less time.

DataverseNO is not a separate corporate body, but is owned by and part of UiT The Arctic University of Norway (see R0). This is the reason why there is no formal Service Level Agreement for the operation of DataverseNO, as the institution does not sign contracts with itself, but the service is run within the same framework as for services delivered for external clients, at the Standard Service Level as listed below.

Time to error solution:
The time to error solution is the time passed from when an error is reported until it is corrected and a solution is reported back to the reporter. Time to error solution is defined within normal working hours. Time to error solution can be longer if a third party vendor is involved in the work to resolve the problem.

Standard Service Level and Time to Error solution (TE):
Criticality: The entire service is down, or the error inflicts on the entire service – TE: 8 hours
Criticality: The error has consequences for all users within one customer or inflicts on a critical service within the customer – TE: 8 hours
Criticality: The error inflicts on a limited number of users – TE: 16 hours

All systems (included DataverseNO) and services delivered by the UiT IT department are subject to risk and vulnerability analysis at implementation, at start up, and at regular intervals throughout the lifetime of the systems and services. UiT (including the IT department) has a management system according to ISO27001 [4], and the risk assessments are based on ISO27005 [5] through guidelines and templates developed by UNINETT [6]. In addition, the IT department has an internal quality control system, The Quality Handbook [7], that is largely based on ISO9000 and some NS-EN-standards (standard developed in Europe (CEN) and then set as Norwegian Standard). Due to some overlap between ISO27001/ISO27005 and the Quality Handbook there is an ongoing process at the IT department to align the UiT policies further with the Information Technology Infrastructure Library (ITIL) [8]] in order to deliver the best quality services possible.

The risk management of UiTs IT systems, including DataverseNO, is described in the Information Security Management System [9]. This system consists of a governing, an implementing and a controlling part, and constitutes UiT's overall approach to information security, by securing the confidentiality, integrity and availability of the information.

The Dataverse application provides MD5 checksums [10] to ensure correctness over time. Furthermore, the transfer of data from old to new storage systems includes checks for bit-correctness of all data.

The disk system health is monitored through common vendor-provided monitoring systems automatically failing out malfunctioning disks, and continuous operation is ensured by standard RAID setups. The storage systems are renewed every 6-8 years, which minimizes the risk for long-term deterioration of storage media.

The operations and services of the UiT IT department are based on regular reviews and checks for compliance with the Quality Handbook (Kvalitetshåndboka) and the Information Security Management System Policy for UiT [11].

References:

[1] Quality management standards (Norwegian only):
https://www.standard.no/Global/PDF/Kvalitet/HandoutA4_OversiktKvalitetsledelse_2018-04_web.pdf

[2] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[3] DataverseNO Preservation Policy (see section on Technological Sustainability, Security, and Disaster Recovery):
https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[4] ISO27001 – Information security management systems: https://www.iso.org/isoiec-27001-information-security.html

[5] ISO27005 – Information technology - Security techniques - Information security risk management:
https://www.iso.org/standard/75281.html

[6] UNINETT Risk Management: https://www.uninett.no/infosikkerhet/risiko-og-s%C3%A5rbarhetsvurderinger-ros

[7] Quality Handbook (Kvalitetshåndboka), only in Norwegian: Can be obtained upon request

[8] ITIL – IT Service Management: https://www.axelos.com/best-practice-solutions/itil

[9] Informasjonssikkerhet ved UiT (Information security at UiT) only in Norwegian:
https://uit.no/om/enhet/artikkel?p_document_id=602863&p_dimension_id=88219

[10] Checksum (MD5): https://en.wikipedia.org/wiki/MD5

[11] Information Security Management System Policy for UiT (Styringssystem for informasjonssikkerhet), only in Norwegian: https://uit.no/Content/409330/Styringssystem-07012015-endelig.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

Preservation Plan

DataverseNO commits to facilitate that published data remain accessible and (re)usable in a long-term perspective. The DataverseNO Preservation Policy [1] describes what challenges DataverseNO faces in long-term preservation, the approaches taken, and the commitments given by DataverseNO to address the challenges to long-term preservation of data submitted to the repository. The organization of the policy reflects the seven attributes of a trusted digital repository, as defined by a de facto standard of the digital preservation community [2]:

- OAIS compliance
- Administrative responsibility
- Organizational viability
- Financial and organizational sustainability
- Technological and procedural suitability
- Systems security and disaster recovery
- Procedural accountability

The implementation of the DataverseNO Preservation Policy is described in the DataverseNO Preservation Plan [3], which is organized according to the recommendations in Becker et al. 2009 [4].

Preservation Strategies and Preservation Levels

DataverseNO employs four major preservation strategies to the digital assets stored in the repository, as described in detail in the DataverseNO Preservation Policy: bit stream copying, fixity checking, normalization, and format migration. (Bit stream copying and fixity checking together form bit-level preservation.) These preservation strategies are applied at three levels of preservation according to the type of file format the digital objects to be preserved are represented in. The preservations levels, the access goals for each object group, and the success measures for each access goal are clearly described in the DataverseNO Preservation Policy:

Preservation Level 1:

- Object Group: All objects.

- Applied preservation strategies: Bit Stream Copying, Fixity Checking.

- Access Goals: Authorized users can access copies of the object in the same format it was originally in the last published version. Preservation at level 1 does not ensure that files are accessible in the same software used at time of access.

- Success Measures: Checksum at time of original processing is the same as at time of future access.

Preservation Level 2:

- Object Group: All objects.

- Applied preservation strategies: Normalization.

- Access Goals: Authorized users can get a copy of the data and documentation files that make up a Dataset in a preferred file format that was current at time of capture or ingest, with significant characteristics of the original as represented in the last published version reasonably intact.

- Success Measures: The normalized versions of all files that make up a Dataset have checksums that are identical to the ones derived at the time of normalization.

Preservation Level 3:

- Object Group: Objects in preferred file format(s).

- Applied preservation strategies: Format Migration.

- Access Goals: Authorized users can access the resource in file formats that are current at the time of access. Files may not correspond one-to-one with the original files, but the significant characteristics of the original resource as represented in the last published version will be reasonably intact.

- Success Measures: The migrated version of the resource retains as many of the significant characteristics of the obsolete version as is practical. Migrated versions of the original are usable in software common at time of access. Migrated versions of all files have future checksums that are identical to the ones derived at the time of migration.

The processes and infrastructure involved in each preservation strategy are described in detail in the DataverseNO Preservation Plan; cf. the sections "Process Characteristics" and "Infrastructure Characteristics".

Deposit Requirements and Transfer of Custody

According to the DataverseNO Accession Policy [5], the DataverseNO Deposit Agreement [6], and the DataverseNO

Deposit Guidelines [7], Datasets to be published in DataverseNO must fulfil a number of requirements to support long-term preservation, including the following:

- Each Dataset must include metadata and a ReadMe file containing information required to identify, verify, interpret, and use the data.
- Whenever possible, Data Files have to be in preferred file formats suited for long-term preservation as advised on by the repository.
- The Depositor grants DataverseNO the right to convert the deposited Data Files and/or Metadata Files to any medium or format and make multiple copies of the deposited Dataset for the purposes of security, back-up, and preservation.
- For the same or other purposes, the Depositor grants DataverseNO the right to make changes to Descriptive Metadata.
- The Depositor grants DataverseNO the non-exclusive right to reproduce, translate, and distribute the Dataset in any format or medium worldwide and royalty-free, including, but not limited to, publication over the Internet.

DataverseNO provides information about preferred file formats in the DataverseNO Deposit Guidelines as well as through advice during data curation.

The DataverseNO Deposit Agreement clearly communicates to the Depositor that DataverseNO requires certain permissions and warrants, including transfer of custody of the Datasets to properly administer DataverseNO and preserve the contents for future use.

Roles and Responsibilities
The DataverseNO Preservation Policy describes the roles and responsibilities that the different stakeholders in DataverseNO have in the development, operation, and maintenance of the DataverseNO Preservation Program as follows:

- Depositor: The role played by those persons or client systems that provide the information to be preserved. Depositors are members of the Designated Community of DataverseNO. Depositors are responsible for complying with established deposit requirements and working with the Research Data Service staff of the repository to ensure a successful data deposit, as well as assist.
- Curator: Research Data Service staff employed at the owner institution and the partner institutions of DataverseNO taking care of ongoing curation of specific collections. Curators check deposited Datasets for compliance with the DataverseNO policies and guidelines, and provide guidance to Depositors on how to adjust deposited Dataset to become compliant with these policies and guidelines before the Datasets are published by the responsible curator. Curators also take care of specific long-term preservation operations as specified by the repository management and the collection management.
- Collection Management: Research Data Service staff employed at the owner institution and the partner institutions of DataverseNO taking care of the management and operation of their collection. The collection management are responsible for specific long-term preservation operations as described in this Preservation Policy, and further specified by the repository management.

- Repository Management: Research Data Service staff employed at the owner institution of DataverseNO taking care of the management and operation of the DataverseNO repository. The repository management take care of the establishment, review, revision, and implementation of the DataverseNO preservation policy, including the long-term preservation operations not delegated to the collection management.
- Advisory Committee: The advisory committee for DataverseNO, and the advisory committees for collections within DataverseNO give advice to the repository and collection management as well as to the Board of DataverseNO on any aspects of Digital Preservation relevant for the repository.
- Board: The Board of DataverseNO have the overall responsibility for all aspects of the DataverseNO preservation policy, and for developing and keeping DataverseNO abreast of the challenges of Digital Preservation in a long-term perspective.

The DataverseNO Preservation Policy describes the concrete tasks that are assigned to the different stakeholder groups in implementing the current preservation plan for the repository.

Preservation Action Plan

To ensure that actions relevant to long-term preservation are taken DataverseNO has – as part of the DataverseNO Preservation Plan – defined a preservation action plan containing concrete actions to be undertaken by the responsible stakeholders and applying the procedures as defined in the DataverseNO Preservation Plan. For each action, the Preservation Action Plan lists the preservation issue, the preservation strategy, the preservation action, the asset group(s), and the time frame applying to the action.

References:

[1] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[2] RLG/OCLC Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities, 2002. https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

[3] DataverseNO Preservation Plan:
https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/preservation-plan/

[4] Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., & Hofman, H. (2009). Systematic planning for Digital Preservation: evaluating potential strategies and building preservation plans. International Journal on Digital Libraries, 10(4), 133–157. https://doi.org/10.1007/s00799-009-0057-1

[5] DataverseNO Accession Policy: https://site.uit.no/dataverseno/about/policy-framework/accession-policy/

[6] DataverseNO Deposit Agreement: https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/

[7] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data and metadata quality

In order for the Designated Community to be able to assess the substantive quality of the data published in the repository, DataverseNO provides documentation of the data in two main ways: On deposit, metadata must be entered into metadata schemas in the repository software (Dataverse), and a ReadMe file must be uploaded together with the data file(s). The repository strives to provide enough domain-specific information about the data such that the Designated Community can assess the substantive quality of the data. However, the generic nature of the DataverseNO repository puts some limitations on the granularity of the provided domain-specific metadata schemas. To compensate for such limitations, domain-specific information is provided in the mandatory ReadMe file.

The deposited ReadMe file must give a description of how to interpret, understand and (re)use the dataset, including a statement of the creation and completeness, or the limitations, of the dataset. The remaining content of the ReadMe file varies according to type of data that are deposited. The DataverseNO Deposit Guidelines [1] give some recommendations for ReadMe files for two common types of data, tabular data and computer scripts. If needed, advice on other types of data is given to depositors on request before data deposit and/or as feedback during the curational review of datasets submitted for publication. In addition, we recommend depositors to insert important parts of the ReadMe file into the Description field in the Citation Metadata of the repository software in order to increase the searchability of the dataset.

The metadata entered into and stored in Dataverse on deposit are standard-compliant metadata to ensure they can be mapped easily to standard metadata schemas and be exported into JSON format (XML for tabular file metadata) for preservation and interoperability. The metadata schemas in Dataverse employ a number of metadata standards from several academic disciplines [2]. All of these metadata schemas are available in all collections of DataverseNO. Citation metadata fields that are mandatory or recommended by DataCite are mandatory in all DataverseNO collections. As the institutional collections within DataverseNO as well as the top-level of DataverseNO accept data from all academic disciplines, which metadata fields are mandatory and which are recommended varies from subject to subject. Special collections within DataverseNO have their own rules for the mandatoriness of, and the recommendations for, domain-specific metadata fields. Depositors are recommended to add domain-specific metadata in the metadata schemas that are applicable; cf. DataverseNO Deposit Guidelines [1].

To ensure compliance with the DataverseNO Accession Policy [3], and the DataverseNO Deposit Guidelines [1], regarding completeness, organization and documentation of the data, each dataset is curated by Research Data Service staff before publication. The curation process ensures that datasets are furnished with relevant information that allows for well-informed reuse of the data. If a dataset does not comply with the DataverseNO Accession Policy and the DataverseNO Deposit Guidelines the curator communicates with the depositor to request necessary changes before the dataset can be published. Changes made to data file(s) and/or metadata after initial publication result in a new version of the dataset and are subject to a new round of curational review before the new version can be published. See also R7, R8, and R12.

Through discussions within the Network of Expertise among the curators, as well as in the DataverseNO Advisory Committee, DataverseNO makes a continuous effort to ensure consistency in both generic and domain-specific metadata across the different collections of the repository.

The quality of data curation in DataverseNO relies on the subject-expertise and the research data management expertise of Research Data Service staff at the DataverseNO owner institution and the DataverseNO partner institutions. This expertise, as well as the roles and responsibilities, are described in R5 and R6. The Research Data Service staff curating the different collections within DataverseNO are all highly educated and trained within the research disciplines represented by the datasets deposited into DataverseNO. The Research Data Service staff are also trained in research data management support, and they are in continuous dialog with the user groups of DataverseNO. Furthermore, DataverseNO Research Data Service staff have access to top-level expertise in subject-related issues and issues on research data management, both through their own networks and through training and advice provided by UiT The Arctic University (owner of DataverseNO). The management and Research Data Service staff of DataverseNO are closely following the development of domain-specific metadata standards as well as other international standards for research data management, such as Domain Data Protocols (DDPs) [4]. This framework aims to support research communities in setting up protocols for the collection and management of data within specified disciplinary domains and research communities.

Automated assessment of metadata

Some metadata fields in Dataverse automatically assess the adherence to the relevant schema. This is e.g. true for the format of dates, the names of language etc. Furthermore, the values of some of the metadata fields (where possible) are generated automatically by the system. This includes the name of the depositor, which is retrieved from the LDAP log-in information, and the deposit date. Some other fields are pre-populated in the metadata templates that are applied for the individual collections. Metadata may be provided both at dataset level and at file level. This is also true for provenance information, which may be provided in two forms: as a provenance file in JSON format and following W3C standards, and/or as a free-text provenance description.

Feedback from Designated Community

The landing pages of each dataset published in DataverseNO has feedback options for the user community to use for comments to the depositor. The default option is to use the contact button to send a question, request or feedback to the contact person for the dataset. DataverseNO does currently not provide end users the possibility to enter annotations or public comments to the datasets, other than by using general web annotation tools like hypothes.is [5].

Citation to related work

The citation metadata schema provides metadata fields for related datasets, related publications, and related other materials. DataverseNO will also benefit from the cooperation between DataCite and Crossref in the Framework for Scholarly Link eXchange (Scholix), that will provide interlinking between datasets and publications based on the datasets [6]. Furthermore, in a future version of Dataverse, planned to be released in 2019, the repository software will implement Make Data Count recommendations and report standardized usage metrics [7]. DataverseNO will use the output from these services (e.g., FAIR usage statistics) to provide additional links to related works and usage metrics together with datasets.

References:

[1] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

[2] Metadata References in the appendix to the Dataverse User Guide:

http://guides.dataverse.org/en/latest/user/appendix.html

[3] DataverseNO Accession Policy: https://site.uit.no/dataverseno/about/policy-framework/accession-policy/

[4] Science Europe: Presenting a Framework for Discipline-specific Research Data Management:

https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

[5] Web annotation tool hypothes.is: https://web.hypothes.is/

[6] Framework for Scholarly Link eXchange (Scholix): http://www.scholix.org/

[7] Make Data Count (MDC) project: https://makedatacount.org/

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

The archiving workflow from deposit to dissemination is described in the DataverseNO Deposit Guidelines (aimed at depositors) [1], and the DataverseNO Curator Guidelines (aimed at Research Data Service staff) [2]. The archiving workflow consists of the following steps:

Step 1

The depositor creates a dataset by filling in mandatory and additional metadata, usually using a metadata template, and by uploading one or more data files in addition to a ReadMe file containing documentation of the dataset. Upon creation, the dataset is not published yet, but only saved as a draft. This draft may be changed or deleted. Upon creation, a draft dataset and all of its file are assigned each their valid DOI. Though valid, while in draft state, these DOIs are not activated and resolvable until the dataset is published.

Step 2

When ready to publish, the depositor submits the dataset (draft) for review.

Step 3

The submitted dataset is reviewed by Research Data Service staff.

Step 4a

If the dataset complies with the DataverseNO Deposit Guidelines, it is published by Research Data Service staff. The dataset and file DOIs are activated and become resolvable, and the workflow has reached the dissemination stage.

Step 4b

If the dataset does not comply with the DataverseNO Deposit Guidelines, it is returned to the depositor with comments on necessary changes.

Step 5

The depositor makes the necessary changes.

Step 6

The depositor submits the dataset (draft) for another review.

Step 7

The dataset is reviewed again by Research Data Service staff, followed by (a) new round(s) of step (4a) or steps (4b) to (6) and 5), until the dataset is ready for publication.

If the depositor does not agree to make necessary changes, the curator addresses the problem by raising the issue within the curator community of DataverseNO to reach a conclusion. If the reached conclusion is not accepted by the depositor, the issue will be raised to the Board of DataverseNO, for a final decision.

A published dataset may be changed. All changes result in a new version of the dataset. Every new version has to be submitted for review before it can be published; see steps (2-7) above.

The handling of data is clearly described and communicated to depositors and users through several policies and guidelines:

The DataverseNO Accession Policy [3] and the DataverseNO Deposit Guidelines describe the criteria and procedures for appraisal and selection of data to be deposited in DataverseNO, how the data should be prepared for depositing, and how deposited data will be disseminated. Data that do not fall within the mission/collection profile as described in the DataverseNO Accession Policy are refused. The refusal of data is communicated to the depositor by Research Data Service staff by email, as described in the DataverseNO Curator Guidelines.

The DataverseNO Curator Guidelines describe in detail how submitted datasets should by reviewed by Research Data

Service staff.

The DataverseNO Preservation Policy [4] describes how deposited datasets are handled for long-term preservation.

The DataverseNO Deposit Agreement [5] describes the transfer of custody and rights from the depositor to DataverseNO to handle the deposited data and metadata.

DataverseNO is a repository for open data. Sensitive data are not accepted for publication. User account information about depositors is handled by Feide, the Norwegian federated log in service, and thus compliant with the Norwegian Act relating to the Processing of Personal Data regulations [6].

Before publishing, deposited datasets are curated as described in the DataverseNO Deposit Guidelines (aimed at depositors) and the DataverseNO Curator Guidelines (aimed at Research Data Service staff). The control of deposited data is regulated through the DataverseNO Accession Policy and the DataverseNO Deposit Agreement. In addition, DataCite will perform an automatic compliance control of the core metadata elements (as defined in the DataCite Metadata Schema [7]), before minting a DOI to a dataset. Also, the repository application, Dataverse, provides automatic output checking of ingested data files by assigning a checksum (MD5) [8] to all files, and a Universal Numerical Fingerprint (UNF) [9], – a unique signature of the semantic content of tabular digital objects.

The roles and responsibilities regarding decision handling within the workflows are described in all relevant DataverseNO policies. As a general rule, everyday workflow decisions are handled by Research Data Service staff of the individual collection in question, whereas decisions regarding more substantial matters are handled by the responsible person(s) or bodies described in the relevant DataverseNO policies.

Changes of workflows have to be sanctioned by changes in the relevant DataverseNO policies. Each policy includes an overview of the policy document's version history.

References:
[1] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/
[2] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/
[3] DataverseNO Accession Policy: https://site.uit.no/dataverseno/about/policy-framework/accession-policy/
[4] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/
[5] DataverseNO Deposit Agreement: https://site.uit.no/dataverseno/about/policy-framework/deposit-agreement/
[6] The Norwegian Act relating to the Processing of Personal Data regulations:
https://www.datatilsynet.no/en/regulations-and-tools/regulations-and-decisions/norwegian-privacy-law/personal-data-act/
[7] DataCite Metadata Schema: https://schema.datacite.org/
[8] Checksum (MD5): https://en.wikipedia.org/wiki/MD5
[9] Universal Numerical Fingerprint (UNF): http://guides.dataverse.org/en/latest/developers/unf/index.html

Reviewer Entry
Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

DataverseNO has a basic search window, as well as an advanced search option. Users can search the entire contents of the DataverseNO, including individual collections, datasets, and files. The search window is available at any level and individual collections in DataverseNO. The search window accepts search terms, queries, or exact phrases (in quotations). The Advanced Search gives the ability to enter search terms for individual collections, dataset metadata (citation metadata and domain-specific metadata), and file-level metadata. Users may also search for variable level names and labels in tabular data files.

DataverseNO is committed to using standard-compliant metadata to ensure that metadata can be mapped easily to a selection of standard metadata schemas. The DataverseNO metadata schemas follow the DataCite metadata requirements necessary to be assigned DOIs [1]. A DOI is automatically allocated via DataCite for each dataset and for each file contained in a dataset.

For each dataset as well as for each file contained in a dataset, the system automatically generates a recommended reference, according to a standard syntax, including the persistent DOI url, and the version number of the dataset. The reference is presented at the top of the landing page of each dataset, and is also available in different formats (XML, RIS,

BibTex). Research Data Service staff at UiT The Arctic University of Norway (owner of DataverseNO) is closely following the development of data citation standards, e.g. through the work by FORCE11 [2]. Furthermore, Research Data Service staff curating TROLLing are contributing to the development of data citations principles for linguistic data by participating in the RDA Linguistics Data Interest Group [3].

Metadata from DataverseNO are exportable to multiple standard formats for preservation and interoperability: Dublin Core [4], DDI [5] and JSON format (XML for tabular file metadata) [6]. Schema.org-compliant discovery metadata are available at the landing page of each dataset [7].

To enhance discoverability of content, DataverseNO also supports OAI harvesting through OAI-PMH [8] [9], and the URL for harvesting is published openly at the site info.dataverse.no [10]. DataverseNO may be harvested as a whole, as well as at the level of individual collections. A near future version of Dataverse will offer metadata in DataCite XML format that is compliant with OpenAIRE [11].

DataverseNO is registered in re3data.org [12]. The metadata of DataverseNO records are indexed/harvested and searchable in a number of discovery services, including DataCite [13], Ex Libris Primo Central Index [14], Bielefeld Academic Search Engine (BASE) [15], and EUDAT B2FIND [16]. Some of the domain-specific collections are harvested by repositories / discovery services targeted toward the relevant researcher communities. TROLLing is harvested by the CLARIN Virtual Language Observatory (VLO) [17], and the UiT Node of the Norwegian Marine Data Centre (NMDC) is harvested by the NMDC repository [18].

References:
[1] DataCite Metadata Schema: http://schema.datacite.org
[2] FORCE11: https://www.force11.org/
[3] Linguistics Data Interest Group: https://rd-alliance.org/groups/linguistics-data-ig
[4] Dublin Core: http://dublincore.org/documents/dces/
[5] DDI (Data Documentation Initiative): http://www.ddialliance.org/Specification/
[6] JSON: https://www.json.org/
[7] Schema.org: http://schema.org/docs/datamodel.html
[8] Dataverse Metadata References: http://guides.dataverse.org/en/latest/user/appendix.html
[9] OAI-PMH: https://www.openarchives.org/pmh/
[10] About DataverseNO: http://site.uit.no/dataverseno/about/
[11] OpenAIRE Guidelines for Data Archives: https://guidelines.openaire.eu/en/latest/data/index.html
[12] DataverseNO record in re3data.org: http://doi.org/10.17616/R3TV17
[13] DataCite search and disseminating service: https://www.datacite.org/search.html
[14] Ex Libris Primo Central Index: http://www.exlibrisgroup.com/products/primo-library-discovery/content-index/
[15] Bielefeld Academic Search Engine (BASE): https://www.base-search.net/
[16] EUDAT B2FIND: http://b2find.eudat.eu/
[17] CLARIN Virtual Language Observatory (VLO): https://vlo.clarin.eu/
[18] Norwegian Marine Data Centre (NMDC): https://nmdc.no/nmdc

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

DataverseNO takes a number of measures to enable long-term reuse of data published in all the collections of the repository.

Required metadata
The general metadata requirements for data to be published in DataverseNO are described in the DataverseNO Accession Policy [1]. Data must be deposited into DataverseNO with descriptive metadata to enable discovery and reuse of the datasets, as described in the DataverseNO Deposit Guidelines [2]. DataverseNO requires and provides documentation of the data in two main ways: On deposit, metadata must be entered into the repository software (Dataverse), and a ReadMe file must be uploaded together with the data file(s). See also R11.

The repository strives to provide enough domain-specific information about the data in order for the Designated Community to understand the data. However, the generic nature of the DataverseNO repository puts some limitations on

the granularity of the provided domain-specific metadata schemas. To compensate for such limitations, domain-specific information is provided in the mandatory ReadMe file.

The deposited ReadMe file must give a description of how to interpret, understand and (re)use the dataset, including a statement of the creation and completeness, or the limitations, of the dataset. The remaining content of the ReadMe file varies according to type of data that are deposited. For details, see R11.

The metadata entered into and stored in Dataverse on deposit are standard-compliant metadata to ensure they can be mapped easily to standard metadata schemas and be exported into the following formats: Dublin Core, DDI, DataCite 4, JSON, OAI_ORE, OpenAIRE, Schema.org JSON-LD.

In addition to general metadata (e.g. citation metadata), Dataverse provide several domain specific metadata schemas [3]. All of these metadata schemas are available in all collections of DataverseNO. General metadata fields that are mandatory or recommended by DataCite are mandatory in all DataverseNO collections. Special collections within DataverseNO have their own rules for the mandatoriness of, and the recommendations for, domain-specific metadata fields. Depositors are recommended to add domain-specific metadata in the metadata schemas that are applicable; cf. DataverseNO Deposit Guidelines.

Following the FAIR data principles, data in DataverseNO are released with a clear and accessible data usage license. See R2.

File Formats
According to the DataverseNO Accession Policy, the preferred file formats for deposited data in DataverseNO are non-proprietary open source or openly documented formats which are extensively adopted by the designated research community and supported by a wide range of software platforms. These formats are best suited to long-term preservation, and reuse and will receive full digital preservation and curation support. In case the original files are not in preferred format(s), preferred format(s) of the data must be provided in addition to the original file format(s). If data cannot be stored in a preferred file format, they can still be published in their original format, but in that case, DataverseNO does not commit to preserve the data in the long term. DataverseNO provides information about preferred file formats in the DataverseNO Deposit Guidelines as well as through advice during data curation. Adherence to preferred file formats is part of the curational review, as described in the DataverseNO Curator Guidelines [4]. File formats not included in the DataverseNO Deposit Guidelines will be assessed during the curation process.

Evolution of File Formats
The DataverseNO Preservation Policy [5] addresses a number of possible challenges to DataverseNO's commitment to ensure long-term access and (re)use of the data published in the repository, among them the evolution of file formats. The preservation policy describes how the evolution of file formats is monitored and acted upon by DataverseNO. In particular, the preservation policy defines several preservations strategies to account for the possible evolution of file formats,

including normalization and format migration. Based on the DataverseNO Preservation Policy, the DataverseNO Preservation Plan [6] describes concrete and measurable actions to overcome, or at least mitigate, the obsolescence of file formats. For details, see R10.

Research Data Service staff closely follow best practice in the field of digital preservation in order to be able to adjust the DataverseNO requirements and to advise depositors on the sustainability of different file formats.

Understandability of the Data

To ensure understandability of the data, each dataset is curated by Research Data Service staff in close collaboration with the author(s) before publication. The objective of data curation is to ensure compliance with the DataverseNO Accession Policy, and the DataverseNO Deposit Guidelines, regarding completeness, organization and documentation of the data. For details about data curation in DataverseNO, see R7, R8, R11, and R12.

The quality of data curation in DataverseNO relies on the subject-expertise and the research data management expertise of Research Data Service staff at the DataverseNO owner institution and the DataverseNO partner institutions. Through discussions within the Network of Expertise among the curators, as well as in the DataverseNO Advisory Committee, DataverseNO makes a continuous effort to ensure consistency in both generic and domain-specific metadata across the different collections of the repository. For more details about the expertise, as well as the roles and responsibilities, see R5 and R6.

References:

[1] DataverseNO Accession Policy: [1] DataverseNO Accession Policy:

https://site.uit.no/dataverseno/about/policy-framework/accession-policy/

[2] DataverseNO Deposit Guidelines: https://site.uit.no/dataverseno/deposit/

[3] Dataverse Metadata References: http://guides.dataverse.org/en/latest/user/appendix.html,

http://guides.dataverse.org/en/4.8.6/admin/metadataexport.html, and

https://dataverse.org/blog/latest-dataverse-update-adds-support-schemaorg

[4] DataverseNO Curator Guidelines: https://site.uit.no/dataverseno/admin-en/curatorguide/

[5] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/

[6] DataverseNO Preservation Plan:

https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/preservation-plan/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

Standards
DataverseNO follows the broad guidance given in the OAIS reference model across the archival process, as described in the section "OAIS compliance" in the DataverseNO Preservation Policy [1]. Changes of any operational and preservation principles for DataverseNO are checked for compatibility with the OAIS reference model, and adapted according to the framework.

The technical infrastructure employed by DataverseNO follows a number of international standards and best practices. Some examples of currently employed standards: The harvesting protocol OAI-PMH is used as a tool to increase visibility and dissemination of content in DataverseNO; the SWORD interoperability standard is used for ingest of structured dataset collections; the Shibboleth/SAML authentication and authorization infrastructure is used as the default for log-in; the industry-standard protocol OAuth 2.0 for authorization is supported and partly implemented; the Schema.org / JSON-LD for structured discovery metadata are implemented and increase the visibility of datasets and support the integration with other services; Docker for operating system level virtualization is being tested as a possible future infrastructure for DataverseNO.

As noted in Requirement R9, the technical infrastructure for the DataverseNO platform is currently running on enterprise class storage and virtualization hardware (VMWare) on a standard CentOS Linux distribution at UiT The Arctic University of Norway (owner of DataverseNO). The infrastructure resides in two datacentres, each in different buildings on campus, where data are replicated to avoid data loss in case of physical threats like fires, floods etc. The VMWare nodes have two power supplies, ups and at least two network cards connected to redundant switches, and the whole operation is monitored continuously with automatic error alerts. Both datacentres are secured with at least two layers of key access doors from public areas, and access is restricted to authorised operational staff.

The development of DataverseNO is an ongoing process strongly influenced also by developments outside DataverseNO, particularly this applies to the system development for Dataverse at Harvard (see below). This means that review of standards and best practices and how they are supported and implemented is done on a more or less continuous basis.

Infrastructure Development
DataverseNO is part of the owner's overall strategy for research data services and is under active development. Currently, the feasibility and evaluation of cloud services for DataverseNO are investigated through national grants applying Docker support for Dataverse. In addition, a future possible DataverseNO infrastructure where both the application and the data are moved into a national or public cloud is actively investigated in cooperation with other national research data services.

System Documentation
The DataverseNO system is run by UiT The Arctic University of Norway (owner of DataverseNO), and system documentation about installation, configuration, integrations and technical operation is kept up to date at a separate SharePoint area within UiTs internal SharePoint domain. Access to this information is restricted to authorized personnel at UiT only. In addition, there is extensive documentation of the Dataverse system provided by the Dataverse Development Community, including Installation Guide, Developer Guide, API Guide, as well as User- and Admin Guide (see below).

Community-Supported Software
The technical repository functions of DataverseNO are provided by the Dataverse software, a widely used open source software developed by an international developer community headed by the Institute for Quantitative Social Science (IQSS) at Harvard University [2]. The current version in use at DataverseNO is 4.15.1. The Dataverse roadmap for new versions is continuously updated. The Dataverse software is hosted on GitHub [3]. Minor releases of the Dataverse software are installed as they become available from the development group at Harvard. Major Dataverse release updates are subject to careful planning and testing before being put into production, in accordance with the Quality Handbook (see R9). DataverseNO continuously evaluates new infrastructure functionalities developed for the Dataverse application, and implement those that are considered useful for the service as a whole.

The system setup is thoroughly documented in the UiT IT department's documentation system (internal) and different

system administrators have performed redeployment of the production platform in order to minimize the vulnerability of the system. In addition to the Dataverse software the DataverseNO platform consists of the PostgreSQL database and a GlassFish application server, as well as standard OS related software. This is all open source software with strong and active community support.

Real-Time Data Streams

For the time being, DataverseNO does not provide real-time to near real-time data streams, but DataverseNO is operated with an around-the-clock connectivity to UiT The Arctic University of Norway (owner of DataverseNO). UiT is the research network hub in northern Norway, and the two UiT datacenters have direct redundant connections to the 100 gigabit/s academic national network backbone operated by UNINETT [4] and thus connectivity into the GEANT network [5].

References:
[1] DataverseNO Preservation Policy: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/
[2] Dataverse: http://dataverse.org
[3] Dataverse on Github: https://github.com/IQSS/dataverse
[4] UNINET: https://www.uninett.no/en
[5] GEANT: https://www.geant.org/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry
Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

DataverseNO is owned by and is part of UiT The Arctic University of Norway, and is not a separate corporate body (see also R0). This is why the security system, security incidents and security handling regarding DataverseNO is an integrated part of the security system, security organization and security administration at UiT. Several of the topics below are described in more detail in R9.

DataverseNO runs on UiT's centralized storage and virtual infrastructure (VMWare). The backup routine builds on a daily backup with a snapshot of the data and the metadata, as well as the whole VMWare server (see also R9). The backup consists of a full snapshot of the server each 90th day followed by a daily incremental snapshot with an integrity check, until the next full backup. In this way, the state of the virtual machine can be restored 90 days back in time, or files / databases can be retrieved 90 days back in time. Recovery time depends on the amount of data. Currently (850 GB), it will probably take up to 1 hour to take a full restore of the server, including the OS-system as well as the application DataverseNO with all the data. A file or partly restore will normally take less time. A detailed time-to-error statement for DataverseNO is presented in R9.

The policy document Information Security Management System for UiT [1] applies to the entire institution and covers detailed operational routines for daily activities and offered services, including DataverseNO. The aim of this policy is to ensure that UiT be a trustworthy institution when it comes to handling of information confidentiality, information integrity and information availability.

Physical infrastructure:
DataverseNO is run on the physical infrastructure for applications and data storage employed at UiT The Arctic University of Norway (owner of DataverseNO). This infrastructure resides in two datacenters, each in different buildings on the UiT main campus in Tromsø, where data is replicated to avoid data loss in case of physical threats like fires, floods etc. Both datacenters are secured with at least two layers of key access doors from public areas, and access is restricted to authorized operational staff. The two VMWare nodes have each two power supplies, UPS and at least two network cards connected to redundant switches, and the whole operation is monitored continuously by a network monitoring system with automatic error alerts. The data storage is backed-up daily with a complete snapshot of the virtual server, making it easier and faster to restore the running environment in case of a server disaster. The back-up has versioning with a file retention time of 90 days. The backed-up data is stored in a separate data hall than the data hall where the production system is running. The two data halls are located in separate buildings, at a distance of 400 meters.

Operational security:

The DataverseNO system runs on a standard, virtual CentOS Linux distribution in VMware. The system is regularly updated as fixes are provided. Minor releases of the Dataverse software are installed as they become available from the development group at Harvard. Major Dataverse release updates are subject to careful planning and testing before being put into production. Administrator access to the DataverseNO virtual server and the VMWare infrastructure is limited to specific networks. The IT department at UiT have monitoring and alarm systems alerting on-duty personnel.

Information security:

DataverseNO complies with the UiT requirements for good computer use practices [1]. UiT has developed extensive technical and administrative procedures to ensure consistent and systematic information security. Good practice requirements include system security requirements, operational requirements and regular auditing and review.
UiT have an appointed CERT (Computer Security Incident Response Team) [2] led by the IT department's information security officer. The purpose of this is to improve the security of UiT's data network, reduce the number of security incidents and the (potential) harm caused, as well as raise awareness of security issues among IT consultants and end users. This includes any incident affecting information security at UiT, incidents that compromise confidentiality and integrity of data, as well as unwanted incidents affecting the availability of data.

As described above and in R9, DataverseNO provides backup storage (located at two data centers) that prevents data loss, and limits the impact of service outages in the case of disasters. Procedures are implemented at UiT The Arctic University of Norway (owner of DataverseNO) to activate crisis teams to deal with system security disasters, see the Quality Handbook (Kvalitetshåndboka) [3] mentioned in Requirement R9.

DataverseNO is identified by the management of UiT The Arctic University of Norway (owner of DataverseNO) as an essential part of UiT's strategy to fulfil the requirements for research data management from national and international funding agencies, as well as from the Ministry of Education and Research of Norway. DataverseNO has already become a core service for UiT researchers and their partners. UiT The Arctic University of Norway (owner of DataverseNO) commits to ensure the proper management and enduring operation of the repository service in accordance with the responsibilities described in the Steering document for DataverseNO [4]. The DataverseNO Preservation Policy describes the procedures for continuity of access and preservation in case of repository closure. See R10.

All systems and services (included DataverseNO) delivered by the UiT IT department are subject to risk and vulnerability analysis at implementation, at start up, and at regular intervals throughout the lifetime of the systems and services. UiT (including the IT department) has a management system in line with ISO27001 [5], and the risk assessments are based on ISO27005 [6] through guidelines and templates developed by UNINETT [7]. See supplementary information in R9. Due to some overlap between ISO27001/ISO27005 and the Quality Handbook there is an ongoing process at the UiT IT department to align the UiT policies further with the Information Technology Infrastructure Library (ITIL) [8] in order to deliver the best quality services possible.

The risk management of UiTs IT systems, including DataverseNO, is described in the Information Security Management System [1]. This system consists of a governing, an implementing and a controlling part, and constitutes UiT's overall approach to information security, by securing the confidentiality, integrity and availability of the information.

References:

[1] Information Security Management System for UiT, only in Norwegian:

https://uit.no/Content/409330/Styringssystem-07012015-endelig.pdf

[2] Computer Security Incident Response Team - CSIRT: https://uit.no/om/orakelet/art?p_document_id=171411

[3] Quality Handbook (Kvalitetshåndboka), only in Norwegian: Can be obtained upon request

[4] Steering document for DataverseNO: https://site.uit.no/dataverseno/about/steering-documents/

[5] ISO27001 – Information security management systems: https://www.iso.org/isoiec-27001-information-security.html

[6] ISO27005 – Information technology - Security techniques - Information security risk management:

https://www.iso.org/standard/75281.html

[7] UNINETT Risk Management: https://www.uninett.no/infosikkerhet/risiko-og-s%C3%A5rbarhetsvurderinger-ros

[8] ITIL – IT Service Management: https://www.axelos.com/best-practice-solutions/itil

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: