Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository DANS: Electronic Archiving SYstem (EASY) complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.
The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on March 28, 2018.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

Assessment Information

Guidelines Version:	2017-2019 | November 10, 2016
Guidelines Information Booklet:	CTS Requirements 2017-2019
All Guidelines Documentation:	Documentation

Repository:	DANS: Electronic Archiving SYstem (EASY)
Seal Acquiry Date:	Mar. 28, 2018

For the latest version of the awarded CoreTrustSeal for this repository: https://www.coretrustseal.org/why-certification/certified-repositories/

Previously Acquired Seals:

Seal date:	November 21, 2013
Guidelines version:	2014-2017 | July 19, 2013
Seal date:	April 12, 2011
Guidelines version:	2010 | June 1, 2010

This repository is owned by:

DANS
Anna van Saksenlaan 51
2593 HW Den Haag
The Netherlands

T +31 70 349 44 50
E info@dans.knaw.nl
W http://www.dans.knaw.nl/

Assessment

0. Context

Applicant Entry

Self-assessment statement:

– Repository Type:

Domain based repository

– Brief Description of the Repository's Designated Community

The repository EASY, the Data Archive of Data Archiving and Networked Services (DANS), is the subject of this self-assessment. The designated community of EASY consists predominantly of scholars in the humanities and social sciences.

– Level of Curation Performed

B. Basic curation – e.g., brief checking, addition of basic metadata or documentation

– *Outsource Partners*. If applicable, please list them

Data storage management has been outsourced to the department ICT Services of the KNAW Head Office (Afdeling ICT Services – I & A Informatisering & Automatisering van het Bureau van de KNAW), Amsterdam. DANS has a Service Level Agreement SLA with I & A, available on request.

This Service Level Agreement covers a broad spectrum of ICT-services (ranging from office IT Services to Data Storage) provided by I & A. Physical data storage has been outsourced in turn by the ICT service provider I & A to

VANCIS Advanced ICT Services, Amsterdam: https://vancis.nl/en . VANCIS has acquired the ISO 27001 (Information Security Management) certification and the ISO 9001 (Quality Management) certification.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. Mission/Scope

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

DANS is the national Dutch organisation for permanent access to digital research data, with a focus on the humanities and the social sciences. The repository, EASY, is its online long-term archiving system. It has been developed for self-archiving either by individual researchers or, increasingly, by institutional depositors.

DANS is an institute of KNAW (Royal Netherlands Academy of Arts and Sciences) and NWO (Netherlands Organisation for Scientific Research) and is governed by the “NWO-KNAW Collaboration Agreement DANS” (Samenwerkingsovereenkomst) between NWO and KNAW, originally from 2005, updated and replaced by a second version in January 2015. This Collaboration Agreement has been signed by the President of KNAW and the Chairman of NWO respectively December 2014 and January 2015. DANS is, administratively speaking, part of KNAW, which is a legal entity. DANS is therefore not a legal entity on its own.

According to this Collaboration Agreement the mission of DANS is “to set up and maintain a national scientific data-infrastructure”. Its main tasks are:

- To disseminate, preserve for the long term and make usable of relevant scientific or scholarly dataset, according to the newest standards.

- To promote the (re-)use of these datasets.

- To collect and make available information on research in the Netherlands.

- To offer outward-oriented services as a national counter.

The "NWO-KNAW Collaboration Agreement DANS" can be found here:

https://dans.knaw.nl/nl/over/organisatie-beleid/informatiemateriaal/SamenwerkingsovereenkomstDANSKNAW2015.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. Licenses

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Both to depositing and using data, agreements apply; respectively the DANS Licence Agreement and the DANS General Conditions of Use. These agreements are based upon the principles of Open Access and the relevant legislation of The Netherlands and the European Union as well as the applicable codes of conduct for scientific research of the Dutch Association of Universities VSNU.

At its website DANS refers to all relevant legal information and its consequences for depositing and using or distributing data, including models of the licence agreement and the general conditions of use as well as help texts: https://dans.knaw.nl/en/about/organisation-and-policy/legal-information .

Licence Agreement

The depositor always enters into a licence agreement with DANS when depositing a dataset in EASY. By accepting the agreement, the depositor agrees with the provisions of the licence agreement that he or she is entering into with DANS as repository.

This agreement contains the mutual rights and obligations that both the depositor and DANS accept with regard to storage and use of the dataset. The most important points are:

- The depositor grants DANS a non-exclusive licence to store his or her dataset and to make it available according to one of the standard access categories indicated by him or her.

- The depositor declares that he or she is the person or that he or she represents the organisation that holds the rights to the dataset and/or are is acting with permission from other possible right holders.

DANS General Conditions of Use

Applying the DANS General Conditions of Use depends on the access category that has been selected for a data file in EASY. Data files with access category Open Access (CC0 Waiver) can be downloaded and used by anyone, without any restriction. The General Conditions are not necessarily applicable to this access category. Data files with access categories Open Access for Registered Users or Restricted Access can only be downloaded and used by registered users. The General Conditions are applicable to these two access categories. When registering, the user accepts the General Conditions.

Registering for EASY

A user has to register once if he or she wishes to use data files in DANS EASY. Registration is not obligatory for data files in the category Open Access (CC0 Waiver). DANS can use some of the registration data for internal DANS user research. By registering the user also accepts the General Conditions. A user also agrees with the fact that with every dataset that he or she downloads, some of these data (name, organisation and position) will be shown in the DANS EASY section that is accessible for registered data users (“activity log”). The user can opt out of this. The DANS privacy regulations (see R4), available via the DANS website, describe how DANS processes personal data.

Non-compliance:

Non-compliance with the DANS General Conditions of Use is dealt with in article 7 of the DANS General Conditions of Use:

a. In the case of non-compliance with one of these Conditions of Use, the use of the dataset must be terminated immediately at the initial demand by DANS. DANS reserves the right, in such an event, to inform the user's employer. In the event of improper use of personal data, DANS also has the right to inform the Dutch Data Protection Authority (Autoriteit Persoonsgegevens).These measures do not prejudice the competence of DANS,in the event of non-compliance or insufficient compliance with these Conditions of Use, to bring an action against the user.

b. The user indemnifies DANS against all claims other parties may bring against DANS as a direct or indirect result of the fact that the user has not or has incompletely taken these Conditions of Use into consideration.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. Continuity of access

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Responsibility

According to article 3 of the Licence Agreement the repository *shall ensure, to the best of its ability and resources, that the deposited dataset is archived in a sustainable manner and remains legible and accessible.* See: https://dans.knaw.nl/en/about/organisation-and-policy/legal-information

Preservation periods are, beyond this, not formally guaranteed in the Licence Agreement. However, continued preservation is guaranteed by the funding organisations of DANS as indicated below.

Continuity:

As stated earlier DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and Netherlands Organisation for Scientific Research (NWO).

In the event of termination of the NWO-KNAW Collaboration Agreement (mentioned under R1) NWO and KNAW will make arrangements regarding the discontinuity or altered continuity of DANS. It is explicitly stated in this Collaboration Agreement that in the latter case KNAW and NWO will take over the responsibility for the data files archived at DANS and store these elsewhere “in the most responsible manner possible and under equivalent technical conditions” (article 10.6 of the Collaboration Agreement).

The present NWO-KNAW Collaboration Agreement has come into force January 2015 and is valid for a period of ten years.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. Confidentiality/Ethics

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Both the DANS Licence Agreement and the DANS General Conditions of Use are based upon the principles of Open Access and the relevant legislation of The Netherlands and the European Union as well as the applicable codes of conduct for scientific research of the Dutch Association of Universities VSNU: https://dans.knaw.nl/en/about/organisation-and-policy/legal-information

Some datasets are ingested on the base of a collective contract with a data supplier. Often these datasets are ingested in another, sometimes, fully automated way (ftp, SWORD or other).

By adhering to the relevant legislation DANS complies with legal and ethical criteria in research. In both the clarifications of the DANS licence agreement and DANS Conditions of Use (available on https://dans.knaw.nl/en/about/organisation-and-policy/legal-information) this is made clear to respectively depositors and users. In article 4.c of the DANS Licence Agreement the depositor as the holder of rights declares: *c. The Depositor declares that the dataset contains no data or other elements that are, either in themselves or in the event of their publication, contrary to Dutch law.*

Data with disclosure risk are managed and stored appropriately. DANS has an internal system of "disclosure risk classification" (internal document "Risicoklassen" available on request). Data requiring the highest degree of protection, mostly datasets containing special and /or sensitive personal data, are stored with the maximum security level. Distribution takes place in an appropriate way and according to the chosen access categories of the licence agreement. For data with high disclosure risk the access category "Restricted Access" is, as a rule, chosen, often even under additional access conditions, imposed by the depositor.

Although EASY is a self-archiving facility, at ingest datasets are checked and reviewed by the data archive staff to be sure they comply with legal and possibly ethical requirements. Ethical requirements are not yet everywhere in place or clearly formulated. This check is part of the workflow for ingesting datasets. When the requirements are not met or anyway not clearly indicated (lacking informed consent forms for example) depositors will be asked to clarify this. Ultimately, when the answers are not satisfactorily, the dataset could be sent back. As a rule, datasets are not anonymised by DANS itself. Anonymised datasets are accepted. For non-compliance with access conditions (including sensitive data) see the answer given under R2.

There is a procedure for handling datasets containing special (or sensitive) personal data. This procedure, to be followed by both depositors and users, enables responsible storage and access according to the legal requirements.

All DANS staff – including guest researchers, trainees et cetera – are obliged to sign the “Declaration of Confidentiality for Employees” (available on request) in which it is stated that he/she will observe and maintain the utmost secrecy with regard to all confidential information that is supplied or will be supplied to him/her by DANS or by persons designated by DANS.

Non-compliance by staff members with the “Declaration of Confidentiality for Employees” is dealt with by either disciplinary measures or reported for prosecution as a criminal act.

The external service provider of the data archive, I&A, is also part of the KNAW. Regarding confidentiality, the Service Level Agreement (SLA) refers to the “Code of Conduct for ICT and Communication Facilities” of the KNAW. Both documents are available on request.

To ensure that the relevant legal knowledge remains up-to-date one member of staff has been assigned the role of Legal Officer with the explicit task to monitor the developments in this field and report possible changes to the director of DANS.

Non-compliance by users with handling data with disclosure rusk are dealt with in article 7 of the DANS General Conditions of Use, in particular article 7.c: *In the case of non-compliance with one of these Conditions of Use, the use of the dataset must be terminated immediately at the initial demand by DANS. DANS reserves the right to, in such an event, to inform the user’s employer. In the event of improper use of personal data, DANS also has the right to inform the Dutch Data Protection Authority (Autoriteit Persoonsgegevens).These measures do not prejudice the competence of DANS, in the event of non-compliance or insufficient compliance with these Conditions of Use, to bring an action against the user.*

Some guidance is given to users on the handling of audio-visual datasets containing personal data: https://dans.knaw.nl/nl/over/organisatie-beleid/juridische-informatie/DANSrapportjuridischeaspectengetuigenverhalen.pdf (only in Dutch). This concentrates on issues like informed consent, access categories, authenticity and integrity.

In the Privacy Regulations of DANS the manner is described in which DANS processes the personal data in its possession. To be found here:

https://dans.knaw.nl/en/about/organisation-and-policy/legal-information/PrivacyreglementDANSUK_april09versie2.82.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. Organizational infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

In the “Collaboration Agreement” (see R1) it is laid down that DANS is funded by NWO and KNAW.

These two organisations, the two major national research organisations of the Netherlands, have agreed to finance DANS for the period 2015 – 2025, with possible ending of the agreement in between with a one-year notice period. Furthermore DANS is engaged permanently in a range of externally funded projects (like EUDAT, EOSC, Parthenos) as well as European data-infrastructures (CESSDA, DARIAH, CLARIN, EHRI, OpenAire). For more information see https://dans.knaw.nl/en/projects These projects are often running over several years.

In “Sharing data together - DANS strategy policy 2015-2020” the mission, target groups and strategy of DANS are described. The full text of the DANS strategy policy is also available in Dutch on the DANS website: https://dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSstrategienota20152020UK.pdf

Sufficient numbers of appropriately qualified staff are available. These cover diverse IT-competences like IT-support both for hardware and software, IT-development, data archival / documentary skills, research and development, communication as well as staff with the necessary background of the disciplines covered by DANS. See: https://dans.knaw.nl/en/about/organisation-and-policy/staff?set_language=en. Regularly updated job descriptions, as a rule based on profiles of functions needed for the repository activities, exist which set out the required qualifications of the digital repository personnel. There is an organisational chart available for the DANS organisation as a whole. There is ample opportunity to update the professional knowledge, both in one of the long-term (inter)national projects, data-infrastructures or in one of the European data infra-structures in which DANS participates.

The organisational chart of DANS can be found here:
https://dans.knaw.nl/en/about/organisation-and-policy/organigram

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. Expert guidance

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The Scientific Advisory Committee (Wetenschappelijke Advies Raad) offers solicited and unsolicited advice to DANS. This Committee consists of disciplinary experts like researchers and scholars in the fields of DANS with a background in IT applications: humanities, archaeology, social sciences: https://dans.knaw.nl/en/about/organisation-and-policy/steering-committe-and-advisory-boards

External advisers are often consulted on varying fields of expertise and specialised subjects, for example on legal matters.

Concerning in-house advice, the “Research and Innovation department” of DANS supports the DANS services. It focuses on innovating how to make digital information available in a sustainable form during the different phases of the research cycle. Research and Innovation collaborates on different research projects with other parties: see the multi-annual research programme (pdf): https://dans.knaw.nl/en/about/research-and-innovation/research-and-innovation

DANS invites visiting fellows to share their knowledge on data management and archiving, and sees it as part of its task to share the newly acquired knowledge with others, for instance by organising workshops: https://dans.knaw.nl/en/about/research-and-innovation/research-and-innovation .

The designated community, indicated under R0, is monitored through substantial contacts, for instance during data acquisition and ingest, in applied research projects, membership of European Research Infrastructures (like CESSDA, DARIAH, CLARIN), pilot studies with data producers, via training & consultancy and by offering discipline-specific services. The pertaining section in the DANS Preservation Strategy (internal, available on request) describes a selection of activities and services for specific communities, whereas the section “Recurring processes” provides an overview of generic processes for monitoring and improving the quality of the data archive.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. Data integrity and authenticity

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Background Information

The data archive of DANS is designed as a long-term preservation repository for the social sciences and the humanities. The data archive provides long-term access to research data to promote reuse for new research and replication of existing research. The FAIR principles (findable, accessible, interoperable and reusable) are leading for the data archive. Translated this means:

- Self-deposit: The data archive must operate in a sustainable way while dealing with the increasing amount of research data. As a principle, the data producer is responsible for successfully documenting and depositing the data to the data archive.

- Long-term preservation: The data archive preserves data for an indefinite time, hence, it must be prepared for migrations and transformations to ensure the readability and usability of the data.

- Trustworthiness: DANS provides sustained access to research data to enable verification and reuse. The data consumers must be able to assess whether the data in the data archive is can be trusted for replicating research or for new research.

Evidence for the answers in this requirement can be found in the DANS document on preservation policy:

https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreservationpolicyUK.pdf)

Integrity

DANS considers the integrity of its holdings as:

- Completeness: no data have been lost (unintentionally). No datasets or files are lost.
- Correctness: no data have been altered. All datasets and files record fixity. All intended changes record new fixity and provenance information.
- The above can be assessed by all stakeholders of the data archive.

DANS facilitates this by:

- Policies, procedures and IT-support for calculating, recording and comparing fixity information.

- Security policy / measures to prevent access / changes by unauthorized users or viruses.

Fixity checks

Fixity in the Ingest stage:

Fixity during deposit is implemented during transfer phase (via HTTPS, SWORD or ftp). After deposit, fixity is calculated and recorded for each file. Potential errors can be reported to the depositor if fixity information was provided with the deposit. Otherwise, the data archive returns the calculated fixity information and asks the depositor to verify these. Virus-scans are performed periodically for ingest by the web interface and standard for all other ingest ways (like SWORD).

Fixity in Archival Storage:

Fixity for the archival storage is implemented on storage- and on object level. Data is stored and backed-up automatically and the media are automatically monitored for block-level integrity. In addition, the Fedora system of EASY automatically records fixity for each object (data and metadata files). Reported errors will be resolved manually.

Fixity for the user:

Fixity information is currently visible after uploading files via the user interface as MDS checksum. In the license agreement and in the DIP a list of all uploaded files is shown containing the SHA-1 checksum for each file.

Completeness of the data and metadata

The complete chain of the data archive’s custody of all datasets is documented through metadata. All actions are explicit, complete, correct and current. If any actions are performed by the data archive which would result in changes to the data files or their organisation, those actions are performed on a copy of the data. The original submission will always be stored exactly as deposited, within a (virtual) folder called ‘original’. The "original" version can therefore always be said to be an integral copy of the version deposited with the data archive.
The original metadata is printed in the licence agreement.

Changes in data and metadata and version control

The data archive distinguishes between two forms of alteration post ingest:

1. New version and therefore creation of a new dataset when there is a change to the data.

2. Minor change: when there is a change to metadata, descriptive documents or supplementary files

When there is a new (version of a) data set, the data archive recreates all descriptive and structural metadata and retains the old file and the previous AIP within the preservation system. The new dataset is assigned a new persistent identifier. This way, the already existing persistent identifier will continue to refer uniquely to the earlier version of the dataset. The new and the previous dataset are cross-referenced in their respective descriptive metadata. Alternatively, when there is a minor change, this change is documented in the administrative metadata; no new persistent identifier is minted. In the case of data conversion to another file format for preservation or access purposes, the data archive maintains the original file(s). The conversion aims to preserve the content of the data, because this is a significant property of the data. Preservation of other aspects, such as the layout of the input format (the “look and feel”) is considered to be of lesser importance. The data archive does not delete data.

Authenticity

Concerning authenticity ("The degree to which a person (or system) regards an object as what it is purported to be. Authenticity is judged on the basis of evidence"), DANS has a provenance document in which the way of publishing data at DANS is described in general. It also contains information on the way DANS deals with mutations and additions of (meta) data. The public document is available at the website:

https://dans.knaw.nl/en/deposit/information-about-depositing-data/dans-provenance-document-uk.pdf

Strategy for data changes

During the ingest of data, the data archive recommends or enforces the depositor to submit its data using preferred - or accepted formats. The list of preferred formats published on the DANS website:

https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf. For more details, see R8. In this way the depositor is responsible for potential transformations, and as such for ensuring the authenticity of the data.

Additionally, during the archival procedure, an archivist may convert files to preferred formats to ensure long-term preservation and accessibility. If files are migrated, the migrated files will be published with the dataset for use. The original files, as deposited, will always be archived with the dataset. Only the approved/curated data and metadata will be published. If the data and metadata required transformations, then the original data are kept, but not published. The depositor will become aware of these changes when the dataset is published.

An archivist may apply minor changes to the metadata or the directory structure. Larger issues will be consulted with the depositor.

Provenance data

The depositor should provide the available documentation about the creation of the data, and how the data can be used. Part of the assessment of the data by the data archive consists of verifying whether the deposited data have not been corrupted, e.g. whether the checksums are correct and whether the files can be opened without errors. It consists furthermore of verifying whether the deposited data is in preferred formats and authentic and thus represents the complete/correct information collected in the project; a typical error is the omission of the codebook for a survey.

After the curation-and/or acceptance of the dataset, the dataset will be sent to the archival storage, where it will be stored in a redundant manner, and where it will be monitored for fixity changes. When unintended changes may appear, the authentic copy will be determined and retrieved and the event will be recorded in the provenance document. The data and metadata will be published.

The user will receive the curated files from the data archive, without transformations. The original files and the provenance document and/or details about the significant properties are currently not provided via the common user interface, but can be provided upon request.

Logging of changes in data and metadata

A data archivist runs through a standard checklist. All actions resulting from this procedure and performed on a dataset by the data archivist are registered in the internal administration. In the field ‘remarks’ the data archivist can document possible additional changes or special issues concerning the data set.

Comparing of essential properties of different versions

The significant properties are not determined or recorded as such; authenticity is ensured by the professional judgement of the data producer and/or data archivist. When data or metadata transformations are needed because of changes in the general policy on data / metadata formats the data archivist will always keep the original data, judge whether the authenticity of the data will be maintained after the transformation and manually document the transformation in the remarks/provenance-field of the dataset. In some cases, a compromise will be made between the authenticity, usability and (financial/technical) costs. Such a compromise will be based on the relative significance of the lost significant properties and the chance that they can still be retrieved from the original data.

Checking identities of depositors

As part of the assessment of the data by the data archive it is verified whether the depositor and/or data producer can be identified, either via federated identity management or via an email-confirmation routine.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Appraisal

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

For DANS it is crucial that research data remain accessible and usable in the long term. This concerns data generated or captured in academic research in any domain, with a strong focus on the humanities and the social sciences: the designated communities of the data archive. There is however no explicit collection policy for those domains. See also R12 on the selection policy. Rather than actively setting out to acquire specific data, the policy of the data archive is to leave it to the researcher and the communities to decide which data are eligible for long-term archiving and access. Some guidelines for users are given in the section “Preparing” of the DANS web page “Information about depositing data” (https://dans.knaw.nl/en/deposit/information-about-depositing-data). This approach is supported by outreach efforts ranging from consultancy, data acquisition staff liaising with specific academic disciplines and a biannual data prize to providing guidelines for the selection of information objects.

As pointed out under R7 – Authenticity, the data archive recommends or enforces the depositor to submit its data using preferred- or accepted data formats. See for further details on how the data archive checks the completeness and understandability of the data ingested R7- Authenticity (in particular provenance).

For understandability see also R14 for the requirements on metadata.

The list of preferred formats is published by DANS: https://dans.knaw.nl/en/deposit/information-about-depositing-data/file-formats?set_language=en.

The following definitions are used:

• Preferred formats are file formats of which DANS is confident that they will offer the best long-term guarantees in terms of usability, accessibility and sustainability. Depositing research data in preferred formats will always be accepted by DANS.

- Acceptable formats are file formats that are widely used in addition to the preferred formats, and which will be reasonably usable, accessible and robust in the long term.

DANS strongly recommends depositors to supply their data in the preferred formats, but acceptable formats will in most cases also be allowed. If data are submitted in other formats than those in the lists of preferred or acceptable formats mentioned depositors should contact DANS. In principle these other formats will not be accepted, unless there are convincing reasons to do this. Long-term preservation cannot be guaranteed then however.

When the data does not meet the required general conditions (see R7), in particular when metadata are not sufficient, informed consent declarations or codebooks are missing or data are incomplete or do not match the description, the data archivist will request the data producer to fix the issues and redeposit, or the data archivist will resolve the issues. For documented evidence on this point see also R12 (instructions).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. Documented storage procedures

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

DANS has two documents about preservation: the Preservation Policy and the Preservation Strategy. The DANS Preservation Policy https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreservationpolicyUK.pdf) details the commitment to support the long-term management of data and also outlines the roles and responsibilities of all those involved in the collection and management of data. This document is referred to for more details. The DANS Preservation Strategy (internal, available on request) is an elaboration of the policy and outlines the framework in which the data archive operates.

The data archive is committed to taking all necessary precautions to ensure the physical safety and security of the data it preserves. This includes a periodical technology vulnerability scan, the SLA with the data storage provider including a confidentiality statement (see Background Information, Outsource Partners), a procedure for file fixity checking as well as the Declaration of Confidentiality for employees and a periodical safety inventory by the KNAW.

As stated in R4 data with disclosure risk are managed and stored appropriately, using the DANS internal system of "disclosure risk classification" (risicoklassen) for data storage (details see under R4).

In essence, the purpose of archival storage is to ensure that what is passed to it from the ingest process remains identical and accessible. In the data archive the archival storage function receives AIPs (Archival Information Packages) from the ingest function and adds them to the permanent storage facility, oversees the management of this storage, including media refreshment and monitoring. This function is also responsible for ensuring that an AIP can be retrieved.

The curate data activity is performed by the data archive and involves the necessary actions to ensure that the data conforms to the collection policy. Such actions may involve format transformations and metadata enrichment, but may also include restoring data from backup (when integrity was damaged). The result is an AIP that will be archived. Curate data is part of the processes "Archive Data" and "Preserve Data". The former will involve curation of data after their deposit, the latter will involve curation after changes in the designated community and/or the introduction of an explicit collection policy.

There is a strategy for backup copies: the data is stored on a dedicated server and every 24 hours tape backups are made which are stored in two physically separate locations. For more details on this see under R15. No data recovery has been necessary so far.

The plans and procedure regarding crises are laid down in the Business Continuity Plan –Crisis management plan DANS 2015 (available on request). The document (in Dutch) includes information on:

- The composition, roles and responsibilities of the crisis management team;

- The protocol when a crisis occurs;

- Information on the susceptibility of the relevant locations for the operation of the data archive.

For checks to ensure consistency across archival copies see R7.

Most measures described above relevant for preservation are part of the instructions for the data archive – see R12 (instructions). Documentation is contained in the AIP.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. Preservation plan

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

As specified in R8 DANS has two levels of preservation, preferred formats and acceptable formats. Supplying data in preferred formats is strongly recommended by DANS, but acceptable formats will in most cases also be allowed. Other formats will not be accepted, in principle, unless there are convincing reasons to do this. Long-term preservation cannot be guaranteed however in the latter case.

The Licence Agreement between repository and depositor provides for all actions necessary to meet the responsibilities. In article 3 it says:

3. The Repository

a. The Repository shall ensure, to the best of its ability and resources, that the deposited dataset is archived in a sustainable manner and remains legible and accessible.

b. The Repository shall, as far as possible, preserve the dataset unchanged in its original software format, taking account of current technology and the costs of implementation. The Repository has the right to modify the format and/or functionality of the dataset if this is necessary in order to facilitate the digital sustainability, distribution or re-use of the dataset.

The handover of responsibilities is clearly stated in the Licence Agreement. This formally takes place after acceptance by DANS of dataset submitted, according to article 10.a of the licence agreement:

10. Term, cancellation and termination of the Agreement
a. This Agreement shall come into effect on the date on which the Repository publishes the dataset (hereafter the date of publication) and shall remain valid for an indefinite period. If the repository decides not to include the

dataset in its data archive, this Agreement is cancelled. The Repository notifies the Depositor of publication or non-inclusion of the dataset in its data archive. Cancellation of this Agreement is subject to a period of notice of six months, and notice shall be given in writing. It is possible to change the agreed access category at any time during the term of the Agreement.

According to the articles 1.b and 1.c DANS is authorised to make copies of the dataset as well as transform, and store the items, as well as provide access to them:

b. The Repository is authorised to include the dataset in its data archive. The Repository shall transfer the content of the dataset to an available carrier, through any method and in any form.

c. The Repository is authorised to make the dataset (or substantial parts thereof) available to third parties by means of on-line transmission. In addition, the Repository has the right, on the instruction of third parties or otherwise, to make a copy of the dataset or to grant third parties permission to download a copy.

Preservation Policy and Preservation Strategy

Both the *DANS Preservation Policy* and the *DANS Preservation Strategy*, referred to in R9, contain the main lines of the DANS preservation plan. The preservation planning is discussed extensively in the DANS Preservation Strategy. The goal of this OAIS function is to ensure that the data in the data archive remain accessible, understandable, and sufficiently usable over the long term. The preservation strategy of the data archive is based upon open and available file formats, data migration and media refreshment. Preservation decisions at the data archive are made within the context of the mission and strategy of the data archive, balancing the constraints of costs, scholarly value, user accessibility, and legal admissibility.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. Data quality

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

During ingest the archivists check the submitted data, metadata and other documentation. This could result in going back to the data depositor/producer to improve the metadata and/or documentation. This could be the case when metadata are not sufficient, informed consent declarations or codebooks are missing or data are incomplete or do not match the description.

Basically, providing Dublin Core metadata is obligatory for all depositions. Some metadata fields are obligatory, while the others are optional. DANS follows as much as possible the specifications of Qualified Dublin Core (see R14 for more details).

More specifically, DANS can keep up with international quality standards in the field of discipline-specific data formats and metadata schemas through intensive cooperation in and with a number of national and international bodies in the field of data archiving and data curation. The designated communities of DANS are disciplines in the humanities and social sciences, in particular also archaeology.

DANS always keeps in line with the standards in these specific fields as much as possible in order to maintain the relevant data and metadata quality for the designated communities. The participation of DANS in several European Research Data Infrastructures is one way of doing this: the CLARIN, CESSDA, DARIAH ERIC’s. These infrastructures represent the interests of large, international, disciplinary communities. CLARIN provides access to digital language data collections across Europe and has introduced a dedicated metadata scheme, called CMDI. Whenever CMDI metadata are available for a dataset the DANS data archive provides them to metadata harvesters, with an extra metadata provider based on the OAI-PMH protocol. Furthermore, DANS has been designated as the service provider for the Netherlands in the CESSDA ERIC, which comprises the European national social sciences data archives. DANS is required to follow the CESSDA requirements on data and metadata. Currently, these are being detailed; DANS has not yet implemented (automated) assessment of social science data according to the relevant schema. Regarding social science data survey data is also the focus of Survey Data Netherlands, a more recent collaboration between CentERdata and DANS, which allows users to browse survey data from various repositories, including the data archive. DANS is also the Dutch coordinator of the DARIAH ERIC, a research infrastructure for the arts and humanities.

Together these infrastructures, with their national branches, enable DANS to stay in close contact with these communities, in particular on the data and metadata formats they prescribe.

Archaeologists in the Netherlands are required to deposit their data in the e-depot for Dutch archaeology, EDNA, which is accommodated at DANS. In order to make submitting data more efficient DANS has implemented the sector's information exchange protocol SIKB0102. DANS has been involved in the development of this protocol. The metadata that the DANS data archive requires is automatically extracted from the "digital packing slip" (in Dutch also known as "pakbon"). Other advantages of adherence to the protocol are the uniform delivery of data and the possibility to link the data to other research information such as archaeological reports.

DANS expects that, as broadening of EASY for other designated communities will take place in the future, more specific discipline-related (domain-dependent) metadata fields will be defined in EASY, notably within the framework of international research infrastructure projects in which DANS is involved.

For a number of datasets in EASY reviews are available. The intention is to introduce a new review system, based on the FAIR principles.

Citations to related works are provided in EASY as standard field in the DIP. EASY automatically generates the citation to the dataset concerned, following the Datacite model. The model for citation can also be found in the DANS General Conditions of Use.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. Workflows

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The work processes have been developed over nearly fifty years. They are described in documents and implemented in the archival workflow, which is based upon the OAIS reference model. Because the EASY system is based upon the principle of self-archiving, the first step in the process is carried out by the data producer. The data producer creates the SIP (Submission Information Package).

DANS provides extensive information for depositors on preparing data for depositing via deposit instructions and background information (see: https://dans.knaw.nl/en/deposit/information-about-depositing-data. In EASY further assistance is provided during the deposit procedure.

The archivist checks the metadata, the privacy clauses, and the file format. When the SIP becomes an AIP (Archival Information Package) preservation procedures ensure its readability over time. All actions related to the preservation of the data are documented in the Provenance document; see https://dans.knaw.nl/en/deposit/information-about-depositing-data/dans-provenance-document-uk.pdf

Data distribution and controlled access are done in an automated way, documented in two documents: the DANS Licence Agreement and the DANS General Conditions of Use (see also R4), https://dans.knaw.nl/en/about/organisation-and-policy/legal-information . On this web page clarifying information can be found on the handling of the data and the restrictions thereof, both for depositors and users.

The curate data activity is performed by the data archive. Further details are described in R9.

For security levels references to the DANS system of “disclosure risk classification” as well as the “Declaration of Confidentiality for Employees” (both more detailed information in R4) are made.

Appraisal and selection

The designated communities are leading in this respect. Data that do not fit in the profiles of the designated communities of DANS are either refused and possibly referred to other data archives in the Netherlands or Europe or, in very rare cases, still accepted because of their academic value.

There is no explicit procedure for the decision process on selection in DANS as the data archive has not, within the limits provided by the designated communities, an explicit selection policy. As stated in R9 it is left to the researcher and the communities to decide which data are eligible for long-term archiving and access, as seen from the perspective of the possible value for the discipline concerned in combination with the suitability of the data formats, the quality of metadata and costs. In collaboration with 4TU.Datacentrum and SURF, DANS has provided the publication “Selection of research data - Guidelines for appraising and selecting research data (Heiko Tjalsma and Jeroen Rombouts eds.)” https://dans.knaw.nl/nl/over/research-en-innovatie/peer-reviewed-publicaties as well as a checklist on selecting research data:
http://www.researchdata.nl/en/services/data-management/selecting-research-data/ .

The types of data managed are the standard types used in the social sciences as well as in the humanities.

Workflows are constantly monitored by the management of the data archive and permanently improved.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. Data discovery and identification

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

As re-using research data is key to DANS's mission, data discovery is supported in various ways. The datasets can always be found in EASY by browsing the various research disciplines or by searching the metadata or by reading the metadata, which is always visible to anyone, irrespective of Open or Restricted Access regimes. DANS stimulates adding as many metadata and other relevant documentation as possible.

DANS Metadata search

The DANS data archive has a search function directly available in the header of the website. See https://easy.dans.knaw.nl/ui/home

Alternatively, an advanced search can be done into target specific Dublin Core metadata fields: https://easy.dans.knaw.nl/ui/advancedsearch . A browse feature is additionally available: https://easy.dans.knaw.nl/ui/browse The results will be lists of datasets, described by metadata.

DANS participates in NARCIS, "the gateway to scholarly information in the Netherlands" (http://www.narcis.nl/?Language=en) to increase data discovery and to relate researchers, research data, and publications. By means of the OAI-PMH protocol (http://www.openarchives.org/pmh) not just NARCIS, but anyone can harvest the metadata in EASY. Datasets of DANS are included in portals of European Data Infrastructures in which DANS is involved, like CESSDA or CLARIN. Some of these are still in a developmental stage.

Furthermore, DANS has opened up EASY for searching the metadata via Google.

Data citations

DANS follows, in the General DANS Conditions of Use, the Datacite guidelines (see R11).

Identifiers

Datasets in the data archive can be identified by three kinds of identifiers:

- internal Fedora identifiers for objects like datasets, files, and folders ?

- external URN (persistent) identifiers for long-term identification of datasets within the national repository infrastructure

- external DOI (persistent) identifiers for citation of datasets

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. Data reuse

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data archive contains descriptive metadata on two levels: the Dataset level and the File Item level. In the instructions for depositing (see https://dans.knaw.nl/en/deposit/information-about-depositing-data) DANS stimulates data producers to document their source material, research methods and publications related to the data.

In these instructions there is special attention for disciplinary differences. The instructions describe what is obligatory and what is (strongly) recommended. The use of DDI metadata (DDI = Data Documentation Initiative, http://www.ddialliance.org/) is obligatory for one special (longitudinal) social science research program funded by NWO.

Besides from this, providing Dublin Core metadata (http://dublincore.org/documents/dcmi-terms/) is obligatory for all depositions. DANS follows as much as possible the specifications of Qualified Dublin Core (http://dublincore.org/documents/dcmi-terms/ . Some metadata fields are obligatory, while the others are optional. DANS has considered to make more fields obligatory, but has decided against it on the grounds that metadata might then become a threshold resulting in researchers not offering their data. As a general policy, the number of obligatory fields is therefore kept as low as possible.

To ensure that data producers observe the description guidelines, the archivists check the submitted metadata (along with the data and other documentation). This often results in contacting the data producer to improve the metadata and/or documentation.

DANS monitors the evolution of the standard formats as used in the social sciences and humanities (included archaeology) in and outside the European Research Data Infrastructures closely (see R11). For possible future migrations special project plans will be developed if the need arises.

Understandability of the data is ensured by the metadata and optional additional (background) documentation of the dataset.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. Technical infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

DANS follows the OAIS reference model across the archival process without significant deviations (http://public.ccsds.org/publications/archive/650x0m2.pdf). There is considerable support for Ingest, Archival storage, Data Management and Access. The Preservation Policy https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreservationpolicyUK.pdf) relates DANS's policy to these functions.

Whereas data curation, for example, is mostly done manually (see Guideline 12), other processes, like making datasets findable, guarding restrictions and distributing the data, are automated. Access to datasets is facilitated by the use of persistent identifiers (R13). A content model has been created for collections, folders, and files.

EASY is based on a Fedora Commons repository. It runs on Red Hat Enterprise Linux. Services to the Designated Community such as deposition and dissemination front-ends are implemented as separate services. For institutional depositors, there is a machine-machine deposit interface, based on the SWORD 2.0 protocol.

EASY development is an ongoing process and also relates to developments outside DANS. This means that standards are reviewed on a more or less permanent basis. DANS has a long-term development plan for EASY, which is the basis for year plans (available on request).

IT infrastructure is a strategic component of DANS to achieve its mission and provide long-term access to digital research data. Digital preservation involves the retention of both the information object and its meaning. It is therefore necessary that preservation techniques can understand and recreate these to ensure the authenticity and accessibility of the data. Out of the main preservation strategies –information migration, technology emulation and encapsulation–the data archive has selected the information migration strategy, because the information content is considered more important ("significant") than its look and feel. Migration focuses on the preservation of the intellectual content. To preserve the data, the data archive migrates the original file format to an open format that is independent of the particular hardware and software that were applied to create them. File formats that are non-proprietary, or are proprietary but in widespread use, will tend to retain the best chance of being readable in the future. Non-standard proprietary formats, used only by a specific software program or specific software version, are likely to present problems for future use.

Community supported software is used as much as possible. Standards for the social science are followed as well as (informal) standards in fields of humanities as archaeology and history. See for preferred formats R8.

Storage Management

This includes activities for storing data, retrieving data and monitoring data, as well as organising and monitoring the redundancy. The data is stored on a dedicated server and every 24 hours tape backups are made which are stored in two physically separate locations. The monitoring of the servers and tape backups is automated and aimed at detecting and resolving errors in the security of the data. Fast storage is used for the operation of the archival system.   The storage array for the original datasets and the tapes are replaced at regular intervals. Access is restricted to engineers who perform a fixed and specified number of maintenance roles. No data recovery has been necessary so far.

Almost all storage is outsourced (as indicated under Background Information, Outsource Partners). The outsource data centres are redundant, secured and well-connected, located in Amsterdam and Almere. Monitoring is used to detect and resolve errors in the security of the data. Fast storage is used for the operation of the archival system.

Technical support is provided for the IT infrastructure by DANS. This ranges from resolving hardware problems of laptops and desktops, to problems with the operating systems or installed applications. In addition, technical support liaises with the external IT providers.

Application support is provided for by DANS internally for the systems developed by DANS. These include mainly EASY, but also the DANS website, the ERP Filemaker database, the PID (URN) server and the Data statistics application. In addition, application support liaises with providers of related IT components.

Functional support for the data archive ensures that the data archive functions as it is supposed to function, and that the functions meet the needs of all users: data managers, partners and designated communities. They coordinate/delegate bugs and changes to application-or technical support and/or liaise with service providers.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. Security

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Security policy

The security policy at DANS is in place to ensure that valuable digital data remains available (availability) and does not become corrupted (integrity) or fall into the wrong hands (confidentiality). If data that is managed by DANS is lost or damaged it might jeopardise confidence in DANS as data custodian. A serious breach of information security could even endanger the continued existence of DANS.

To provide optimal access to the digital objects DANS follows a fixed procedure if EASY is interrupted – whether scheduled or not. When EASY should break down in an unscheduled way, users are informed by a warning page.

Depending on the nature of the problem, DANS can use the audit log to trace back data producers who were depositing data at the moment of the interruption. By contacting them potential data loss can be prevented. In less severe cases of interruption EASY can be accessed in read-only mode, which excludes ingest, but allows for data access including dissemination.

The ICT infrastructure is secured by several security measures and security layers. The EASY servers are physically located in a secured environment with strict access rules. The storage of the EASY data is backed-up daily. The back-up has versioning with a file retention time of 90 days. The back-up is stored within the Netherlands on two different locations.

The plans and procedures regarding crises are laid down in the Business Continuity Plan – Crisis management Plan DANS 2015. For details see under R9.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

17. Comments/feedback

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

0. N/A: Not Applicable.

Self-assessment statement:

None

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: