Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository Mendeley Data complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.
The afore-mentioned repository has therefore acquired the CoreTrustSeal on June 22, 2017.

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: Mendeley Data
Seal Acquiry Date: Jun. 22, 2017

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

Previously Acquired Seals: None

This repository is owned by: **Mendeley Data**

T +447454698869
E mike.jones@mendeley.com
W http://data.mendeley.com/

Assessment

0. Context

Applicant Entry

Self-assessment statement:

1) Repository Type:

Other - Mendeley Data is a generalist open research data repository, applicable to all areas of science. All file formats and types may be uploaded, although, as indicated in further detail in the application, certain file types more suitable for preservation are preferred. Mendeley Data aims to openly interoperate with other repositories and systems, via implementation of open standards and a public API.

2) Brief Description of the Repository's Designated Community:

Mendeley Data's Designated Community is similarly broad, as it may include researchers in any field of science. We aim to support the data needs of both institutional researchers, and authors submitting papers for publication.

3) Level of Curation Performed:

B. Every published dataset is reviewed by a trained Mendeley Data reviewer, to verify it is scientific in nature, does not constitute a previously published article, and as far as possible that it does not contain personally identifiable information (as described in more detail in Guideline 4). In cases where the dataset constitutes valid research data but metadata is insufficient or incomplete, feedback will be provided to the author to encourage them to provide additional descriptive information. As the repository is a generalist one, we cannot provide feedback on or curation of the scientific content of datasets. Additionally, it is at the author's discretion to provide additional description in response to our feedback - we don't currently enforce this.

4) Outsource partners:

1. [DANS-KNAW](#) (DANS) is contracted to provide a long-term archiving solution – specifically DANS supports the following: Continuity, Preservation, Data Integrity, Archival Storage

2. [DataCite](#) – reserves and mints DOIs for datasets – therefore supports the Persistence function)

3. [Amazon Web Services](#) – provides storage, hosting and servers – therefore supports Technical Infrastructure and Security

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:

Mendeley Data is a repository launched by Mendeley (part of Elsevier) in 2015, which has an explicit mission to store and preserve data deposited to our repository, and to make the data accessible to Data Consumers for the long-term. Our mission is set out here.

The mission has received backing from senior internal stakeholders, including the Managing Director of Elsevier Research Products.

Research Data Management is a long-term strategic priority for Elsevier. Elsevier has been active in this space since initiating its Data Linking and Research Elements programmes in 2013. Following these successes, a strategic project was started in 2014 that resulted in the pilot project for Mendeley Data in 2015. Since the full launch in 2016 Mendeley Data is now part of the 5 year operation plan that has been accepted by the Elsevier Board May 2016.

We communicate our commitment to preserve datasets deposited to the repository, to our end users, on our homepage, within FAQs, our mission statement, and our archiving processes page. For more information on our approach to preservation, see R10.

To ensure the long-term archiving of datasets posted to the repository, Elsevier has established a non-exclusive archival service provider relationship with Data Archiving and Networked Services (DANS), of the Netherlands.

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:

When posting a dataset to Mendeley Data, Data Producers must select a licence governing access and usage, from a set of licence options, which include Creative Commons licences, software licences and hardware licences.

It is not possible to submit a dataset without selecting a licence, therefore all datasets have a licence. In order to support the Data Producer's licence selection process, we provide explanatory text for each licence, and a link to the relevant licence provider's webpage for full information.

The licences we currently make available to Data Producers are:

• CC0 1.0

• CC BY 4.0
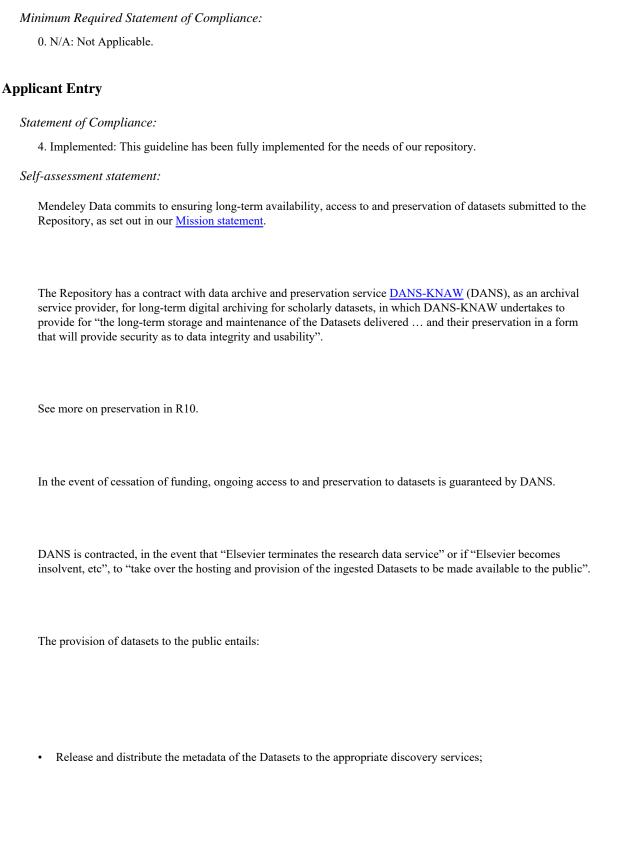
• CC BY NC 3.0

• MIT

• Apache-2.0

- BSD 3-clause

- BSD 2-clause

- GPLv3

- CERN OHL

- TAPR OHL

These licences govern the usage, distribution and author attribution requirements of datasets.

The licence selected by the Data Producer is displayed on the public dataset page, along with a description of the licence's requirements. Therefore Data Consumers accept that they will follow the licence requirements, when they use the data.

If an instance of non-compliance with a licence by a Data Consumer is discovered, then the repository will contact the Data Producer to notify them:

- Which dataset is affected

- The location (e.g. URL) of the offending artifact

Usage of the repository, either as a Data Producer or a Data Consumer, is also governed by our Terms. See requirement R4 for more detail.

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:

Mendeley Data commits to ensuring long-term availability, access to and preservation of datasets submitted to the Repository, as set out in our Mission statement.

The Repository has a contract with data archive and preservation service DANS-KNAW (DANS), as an archival service provider, for long-term digital archiving for scholarly datasets, in which DANS-KNAW undertakes to provide for "the long-term storage and maintenance of the Datasets delivered … and their preservation in a form that will provide security as to data integrity and usability".

See more on preservation in R10.

In the event of cessation of funding, ongoing access to and preservation to datasets is guaranteed by DANS.

DANS is contracted, in the event that "Elsevier terminates the research data service" or if "Elsevier becomes insolvent, etc", to "take over the hosting and provision of the ingested Datasets to be made available to the public".

The provision of datasets to the public entails:

• Release and distribute the metadata of the Datasets to the appropriate discovery services;

- Per Dataset DOI, enable and register a specific URL that links the DOI to the associated Dataset at DANS-KNAW;

- Per Dataset, provide an overview page that lists the metadata of the Dataset, the individual files in the Dataset, and allows downloads of the - metadata and the individual data files, collectively and separately.

Provide access at the following minimal level of functionality:

- The ability to perform searches

- A listing of datasets

- The ability to view metadata and any sub-datasets for a given Dataset DOI

- The ability to download datasets and metadata, consistent with the terms of this Agreement.

-

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:

    3. In progress: We are in the implementation phase.

Self-assessment statement:

As far as possible, Mendeley Data aims to ensure datasets posted to and made available by the repository comply with confidentiality and ethics guidelines.

Terms which govern the usage of the repository, a data producer or data consumer, are published here.

The Terms require (4.4.7) that authors "have obtained all necessary consents"; and that "data is suitably anonymized wherever appropriate". This means that data with disclosure risk must not be shared, except where consent has been given, or should be anonymised.

In our FAQs ("What are the requirements for Mendeley Data datasets?"), we provide guidance to Data Producers, that datasets must not "contain sensitive information (for example, but not limited to: patient details, dates of birth etc.)"

Every published dataset is reviewed by a trained Mendeley Data reviewer. Any datasets found to contain confidential or personally identifiable information (PII), will be taken down per our review and takedown process. We provide a "Notify" mechanism on every dataset page, such that Data Consumers can notify us if inappropriate information is found.

Every Mendeley Data reviewer is trained to examine datasets to establish, to the best extent possible, that the dataset does not contain copyrighted or previously published article papers; is scientific in nature; and does not contain personally identifiable information. The legal responsibility to ensure the latter, per our Terms, is with the Data Producer.

Future features are planned to further increase our support for depositing confidential information securely:

- "Confidential datasets", whereby the data files are not available for download. The author must be contacted to request private sharing of the data.

- A checkbox affirmation, prior to publication, by the Data Producer, to confirm that no confidential information is present in the dataset, and that they have complied with our Terms. This will be a required step to publish the dataset.

Once a dataset has passed review, it is subsequently forwarded to DANS archive.

The DANS review process checks for confidential information as follows: "[data are] processed by a staff member at DANS in accordance with a standard data processing protocol. … On the basis of this protocol, the following types of verification have been performed since the introduction of DANS in 2005:

…Verification of the presence of privacy-sensitive data, both in the files and in the metadata" (DANS contract).

DANS may contact Mendeley Data to request action to be taken, if a file includes personally identifiable information, for instance exact names and exact dates of birth of survey respondents.

This is documented here.

In the case of any personal data being legitimately present within a dataset, these data may only be preserved, stored and used in line with the 8 data protection principles of the UK Data Protection Act (1998):

1 Personal data shall be processed fairly and lawfully.

2 Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

3 Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

4 Personal data shall be accurate and, where necessary, kept up to date.

5 Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

6 Personal data shall be processed in accordance with the rights of data subjects under this Act.

7 Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

8 Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Rev1: depending on the author to approve access requests to confidential datasets is a risk in case the author is no longer able or willing to respond. A contingency plan for such cases would be strongly recommended, e.g. request the author to assign a person or organisation who will take over the responsibility if needed.

Rev2: what happens in case of "orphan datasets" for future access requests?

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:

Research Data Management is a long-term strategic priority for Elsevier. Mendeley Data is part of the 5 year operation plan that has been accepted by the Elsevier Board May 2016 and therefore has funding for that period.

Funding includes provisioning sufficient staff, including a software development team (to continue to evolve the product to meet the designated community's requirements, and for maintenance, bug fixes and operations), a repository manager, and a repository support officer.

[Call out our ability to engage with the community and add features to meet the community's needs and requirements)

In more detail, members of staff assigned to Mendeley Data include:

- Business owner

- Product and software development team: product manager, project manager, business analyst, UX designer, 6-8 software developers, quality assurance engineer.

- Data repository manager role

- Data repository support officer

Mendeley Data dataset reviewers receive internal training to carry out their duties. This includes training in reviewing datasets (as described in more detail in R4), and communicating appropriately with Data Producers.

Expertise and affiliations: Members of the Elsevier research data management team publish and speak widely on research data management (from policy, product and technology points of view), at industry forums such as Research Data Alliance and SciDataCon.

Department members currently co-chair the following Research Data Alliance groups:

- Data Discovery Paradigms Interest Group,

- Data Description Registry Interoperability (DDRI) Working Group,

- RDA/WDS Scholarly Link Exchange (Scholix) Working Group,

- RDA/WDS Publishing Data Services Working Group.

Elsevier is affiliated to FAIR principles (signatory), FORCE11 Joint declaration of Data Citation Principles, ICSU World Data System, etc.

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:

Elsevier has convened a Research Data Management Advisory Board, comprising experts with professional RDM roles in institutions, from industry bodies etc. The Board meets regularly to provide strategic input on RDM activities.

Mendeley Data engages with the community to receive expert advice and guidance in the following ways:

- Works with research institutions and publishers as development partners, to develop the proposition and offering in line with commercial needs.

- Actively seeks feedback from end users, via on-website feedback mechanisms, and a user panel.

- Participates in industry conferences and initiatives, including Research Data Alliance, the Force11 Joint Data Citation Principles, Scidatacon, Force11 etc

Here is a list of Elsevier Research Data Management Advisory Board members.

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:

Mendeley Data ensures integrity and authenticity of data posted to the repository, comprising process and technical aspects.

Authenticity:

Depositor identity is verified, in that the Data Producer must have registered a Mendeley account, providing their name and email address.

The dataset thenceforth is linked to the Data Producer's account, and only the Data Producer may edit, manage and publish changes to the dataset.

Once a dataset is published, the dataset version is fixed and immutable. The Data Producer can edit their dataset, but these edits create a new version. Each dataset version is interlinked to the other versions, allowing a Data Consumer easily to see which is the latest version, and to navigate to access other versions. This allows for a dataset to evolve over time (e.g. for longitudinal studies), but for each version to be citable as a fixed record.

The visible version history, along with versioned (and easily visually comparable) contributor details, description and steps to reproduce, and related links, provide an indication of the provenance of the dataset, in terms of origin and subsequent changes.

Links to metadata and to other datasets are maintained through our Related Links, which allows Data Producers to add links to any associated datasets, software, articles, protocols or other entities. These links are version-specific.

To ensure completeness of data and metadata, as described in more detail in R8, required metadata fields must be completed in order to publish the dataset.

Integrity:

When the dataset is created, the data files are stored on Amazon Simple Storage Service (S3), which ensures integrity of files by calculating checksums and immediately repairing corrupted files using redundant data: "Amazon S3 also regularly verifies the integrity of data stored using checksums. If Amazon S3 detects data corruption, it is repaired using redundant data. In addition, Amazon S3 calculates checksums on all network traffic to detect corruption of data packets when storing or retrieving data." ([Source](#))

Furthermore, when following review, a dataset is archived with long-term archiving partner DANS, "DANS-KNAW checks the checksum delivered with a dataset at ingest. DANS-KNAW runs regular checksums and will be able to detect and repair technical corruption within the archival environment."

In addition, DANS may "in rare events [make] modifications to the content of a Dataset in the DANS-KNAW Archive after it has been delivered by ELSEVIER … Authenticity will be kept by recording modifications as provenance metadata. Versions will be kept where modifications to the content cannot be reversed." [DANS contract]

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:

    3. In progress: We are in the implementation phase.

Self-assessment statement:

In order to ensure that datasets are understandable to our Designated Community, certain requirements are placed on the data files and metadata which may be submitted.

We publish a list of preferred formats, to indicate to Data Producers which formats are suitable for long-term preservation, which is available [here](#).

For these preferred formats, our long-term archive partner, DANS, "guarantees long-term usability", while for all formats it "guarantees long-term bit-level preservation" (Contract between DANS-KNAWL and Elsevier - article 3.1 of the Archiving Agreement Elsevier – DANS).

Additional criteria for datasets are as follows:

Datasets must be:

- Scientific in nature

- Research data - rather than the research article, which may have resulted from the research

Datasets must not:

- Have already been published, and therefore not already have a DOI (Because many such files are article paper PDFs, if a Data Producer uploads a PDF, we provide an alert to ensure the Data Producer is aware that research data rather than article PDFs are desired.)

- Contain executable files or archives that are not accompanied by individually detailed file descriptions.

- Contain third-party copyrighted content - the author may upload copyrighted material only if they are the copyright owner or have the copyright owner's permission

- Contain sensitive information (for example, but not limited to: patient details, dates of birth etc.)

The following mandatory metadata fields must be completed in order to publish a dataset via the web interface or API: (therefore completion of mandatory metadata does not require review)

- Dataset title

- Contributor(s) Names and Email addresses

- Subject discipline categories

- Dataset description

- License

The following optional metadata fields may be completed:

- Steps to reproduce - Here the Data Producer may describe how the data was generated, including any protocols, workflows or software used

- Related links - Here the Data Producer may add links to any related datasets, articles, software, protocols etc

- Contributor(s) Affiliation and Contribution

- Data file description - Here the Data Producer may add file-level information describing the file format, how the data file was generated

While the set of mandatory metadata fields may be small, this is necessarily so to be able to serve our broad Designated Community, posting datasets from a wide range of disciplines; a future enhancement is planned such that partner institutions and journals/publishers will be able to specify more detailed metadata templates for datasets submitted.

Appraisal:

In order to ensure, as far as possible, compliance with the above criteria, every dataset submitted to the repository is reviewed by a trained Mendeley Data reviewer.

The reviewer audits the dataset to ensure it complies with the above requirements, so that it will be relevant and understandable to the Designated Community. In addition, reviewers may add curated links to associated peer-reviewed literature; Data Producers and Consumers cannot add these official links.

Datasets which are found to be already published papers, to be non-scientific, or to contain confidential information, will be removed from the public archive, and the Data Producer notified by email so they may post a new dataset or new version. The DOI which was minted for the dataset at publication will now resolve to a web page indicating the reason the dataset has been removed.

Datasets which pass review are subsequently forwarded to DANS to be long-term archived.

The DANS review process checks file formats as follows: "[data are] processed by a staff member at DANS in accordance with a standard data processing protocol. … On the basis of this protocol, the following types of verification have been performed since the introduction of DANS in 2005:

…Verification of the file format. In the future as well, it should still be possible to open and use the data files as well as the documentation files. The verification is performed on the basis of a list of preferred file formats."

At ingest, the following information about data files is captured:

• Data file size

- File format, where it matches a known format, for instance where a visualization is available for this file format (in all cases, the file extension is displayed to the end user)

We do not currently carry out Collection Development for datasets in particular themes, but we plan a future enhancement to enable community members to do so.

The above information is documented here.

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:

Data Producers may create (or in OAIS terms, "Ingest") datasets either via the web interface, or the publicly-documented REST API (documentation available here).

Data Producers may edit and privately share a draft dataset version before publishing, at which point the version becomes fixed, and the dataset is accepted from the Ingest function by the repository, and is assigned to Storage.

After the dataset has been reviewed by the Mendeley Data reviewer (as described in more detail in R9), it is subsequently forwarded to DANS for long-term archiving and preservation.

The dataset Access function is fulfilled by Mendeley Data, which provides dataset record view and file downloads at the dataset URL (resolvable to via the dataset DOI), dataset listings, and search capability.

In addition, DANS provides Access to datasets' metadata records only, via the EASY archive. Mendeley Data datasets in EASY archive are available here.

Our archival storage processes are documented here. These are reviewed and updated each time a change to our model occurs, for instance in response to a new requirement.

Data Producers must register a Mendeley account, and must authenticate via Elsevier's Access & Entitlements service, before being able to post data.

Data is stored with Amazon S3, part of Amazon Web Services (AWS); no data is stored on local Elsevier Servers. Details of AWS security policies can be found here.

Any corruption of datasets creates error logs; and backups are kept to restore data. Automated database backups happen every day whilst online, with a retention period currently set to 7 days. These backups are stored in the Amazon AWS S3 bucket. Amazon S3 synchronously stores "data across multiple facilities … the objects are stored, Amazon S3 maintains their durability by quickly detecting and repairing any lost redundancy."

Point in time recovery is possible to any point within a second. Restores have been tested periodically to provision development and testing environments. S3 bucket is not backed up due to its high data durability policy. Amazon S3 storage is designed to provide 99.999999999% durability of objects over a given year. Please see page 5 here for further information.

In addition, datasets which pass review are archived with DANS ensuring long-term preservation, where they also benefit from redundancy and backups. "DANS-KNAW stores and archives three copies of any dataset received: two at the same location and one at a different location with a different disaster threat".

Furthermore, "DANS-KNAW checks the checksum delivered with a dataset at ingest. DANS-KNAW runs regular checksums and will be able to detect and repair technical corruption within the archival environment."

Mendeley Data uses checksums to verify the datasets and data files received are exactly those which were sent to DANS. Each file in the dataset is checksummed and so is each part of a transmitted deposit.

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:

We indicate to Data Producers and to our community (via our [homepage](), [mission statement](), [FAQs]()) that we undertake to preserve datasets posted to the repository for the long-term.

This preservation function encompasses: taking delivery of the dataset ingested, storing it, and ensuring it is archived, and accessible and usable to Data Consumers.

To ensure long-term availability of the data, DANS is contracted to "maintain the integrity of the Datasets content … and commits to the long-term and storage and maintenance of these materials."

In terms of active preservation of file formats, "DANS-KNAW will employ appropriate technical solutions to adapt to changes in storage or access technology and to otherwise ensure the continued availability (in accordance with this Agreement) of the Datasets."

In order to ensure that this archiving function is fulfilled:

- DANS provides confirmation of receipt

- Mendeley Data will carry out a random check of 25 datasets archived with DANS, twice per year.

- Additionally a future enhancement is planned such that for each dataset published on Mendeley Data and archived with DANS, a link is supplied from the dataset to the metadata record in EASY, so that the existence of the dataset in the EASY archive can be independently verified.

For data files which are in the preferred formats published by DANS, and in turn by ourselves, DANS understakes to ensure they are usable over the long-term, while for all other formats, they will be accessible.

When the dataset is published, the Data Producer affirms via tickbox in the web form, that the dataset version will become fixed and immutable.

According to the Mendeley Data Terms (published here), Data Producers grant Mendeley Data the rights to copy, transform, and store the items, as well as provide access to them: "For Research Data that you make publicly available on the Site, you grant us a perpetual, irrevocable, worldwide, non-exclusive right and license to publish, extract, reformat, adapt, build upon, index, re-distribute, link to and otherwise use all or any part of the Research Data in all forms and media (whether now known or later developed), and to permit others to do so."

The preservation process is publicly documented here.

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:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Mendeley Data aims to facilitate Data Producers providing high quality datasets, and Data Consumers being able to evaluate datasets based on quality indicators, so they may find data useful for their research.

For every dataset submitted, required metadata must be supplied, and file format advice is provided, as indicated in R8. Metadata fields such as title, description, subject discipline category, steps to reproduce (methods), and related links and articles, aim to ensure that there is enough available information about the data such that the Designated Community can assess the quality of the data.

In addition, the Data Producer can provide additional links to related datasets, articles and software, which give extra context for the dataset, both in terms of content, and in terms of potential quality of research. Mendeley Data reviewers additionally add curated links to peer reviewed literature.

As indicated in R8, a future enhancement is planned such that partner institutions and journals/publishers will be able to specify more detailed metadata.

In order to help the Designated Community evaluate the quality of datasets, we provide, where relevant:

• Dataset quality badges: which indicate whether the dataset:

• Has been reviewed by expert reviewers

- Is associated to a peer reviewed published article.

- (Forthcoming) Has been reviewed and approved to appear on an institutional partner showcase

- Has been reviewed and curated (and updated according to feedback by the Data Producer) by a relevant subject matter

- Dataset metrics: which indicate the number of views and downloads the dataset and data files have received

- (Planned future enhancement) Data citations: the number of citations the dataset has received will be displayed on the dataset record webpage.

- (Future enhancement under consideration) Community comments and up/downvotes on datasets; for instance to provide feedback to the Data Producer on how to make the data more useful, or to indicate their experiences and success or otherwise in reproducing the research or reusing the data.

Additionally, in cases of authors posting non-data datasets (e.g. a research paper), we take down and contact to suggest submitting research data

Elsevier participates in the [RDA Data Fitness For Use Working Group](#) in order to contribute to community-approved standards for dataset quality.

Here are links to publicly available listings of Mendeley Data datasets:

- [Mendeley Data - Datasets catalogue](#) - all published and reviewed (i.e. valid) datasets

- [DANS EASY archive](#) - metadata for all published and reviewed (i.e. valid) and archived datasets

- [DataCite dataset index](#) - metadata for all published and reviewed (i.e. valid) datasets

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:

Mendeley Data's workflow for archiving datasets is archived here. Any changes to workflow will be recorded as new versions of this page, and links provided to previous versions.

A summary of the archiving workflow is as follows:

- Once a dataset Submission Information Package has been published, it is reviewed by a trained Mendeley Data admin reviewer

- The reviewer verifies the dataset meets the requirements – both the requirements and the appraisal process are described in R8. As set out in our Terms, which govern their usage of the repository, Data Producers may not post information with disclosure risk, or which has not been anonymised.

- Datasets which do not pass the Mendeley Data review process, are removed from the website, and the author contacted; the dataset webpage now displays the reason the dataset was removed. The dataset DOI will resolve to this webpage.

- We communicate clearly with Data Producers within the dataset creation process, and within the FAQs and [process page] to advise them that their dataset, if approved, will be stored for the long-term in the repository, and additionally archived with DANS.

- Upon approval by Mendeley Data admin reviewer, our system forwards the dataset, via API, to DANS, for long-term archiving.

- The DANS system confirms receipt via API response, or alternatively returns an error, which the Mendeley Data development team will then investigate.

- Following review by the DANS service, the dataset will appear in the DANS EASY dataset archive.

- Mendeley have contracted with DANS to ensure that all published and valid datasets are archived in perpetuity. If in the future, the Mendeley Data site ceases to exist, all archived datasets will still be available in DANS.

- However, while datasets are discoverable in DANS, only the metadata can be seen. The files themselves are currently only viewable within Mendeley Data.

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:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Every public dataset is available via our public dataset listing webpage, and via our public API.

Keyword searching of dataset metadata and file contents by Data Consumers will be supported, to enable effective data discovery. This keyword search capability will be powered by DataSearch, a data search engine developed by Elsevier. Data Consumers will be able to find relevant results from within Mendeley Data, and also other repositories indexed by DataSearch. This capability is expected in the first half of 2017.

Every public dataset is provided with a unique persistent identifier, in the form of a versioned DataCite DOI (draft dataset versions are provided with a reserved DOI, so that the eventual dataset citation can be known to the Data Producer in advance of publication, to enable it to be referenced in a manuscript for example).

Datasets may also be accessed via dataset links on article pages, where the article platform supports this.

In addition, dataset metadata is available in the DataCite registry, and the DANS EASY archive.

A planned future enhancement is to allow links between articles and datasets to be discovered via the Data Literature Interlinking platform, a Scholix-compliant platform.

Metadata can be harvested via standard markup: at the moment various metadata fields are marked up in Dublin Core and Google Science Datasets markup. A planned future enhancement is to add support for W3C DCAT markup. An additional planned future enhancement is to support harvesting via OAI-PMH protocol.

A dataset citation in standard format is provided for each dataset, to facilitate citing 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:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The repository aims to ensure that sufficient metadata in high enough quality are captured to support understanding and (re)use of data.

As described in R8, mandatory and optional metadata are provided by the Data Producer, when publishing a dataset.

To help ensure understandability of the data to consumers, Data Producers must provide a title and description; they are also encouraged to indicate the steps to reproduce the research which led to the data, for instance methods, workflow and/or software used; and to provide links to any software or other datasets used in generating the data, or associated articles.

In addition, Data Producers may provide a description for the individual files within the dataset, for instance describing the contents, related findings, the format, or the processes which led to the individual file.

As a repository, we are exploring ways to encourage Data Producers to provide intelligible and consistent file labeling and contents, possibly including a data dictionary, schema verification, data curation, etc. This is an area for ongoing improvement.

As described in R8, a future enhancement is planned to allow partner research institutions and journals/publishers to provide subject-specific / content-oriented metadata templates for associated datasets.

As also described in R8, preferred file formats advice is provided to Data Producers.

As described in R2, the Data Producer must select an appropriate licence from a pre-defined range of options, so that they can set their desired conditions for reuse of the data.

To address the issue of future evolution of formats, and any future migrations therefore needed, our long-term archive partner DANS undertakes to ensure all files uploaded in preferred formats are usable in perpetuity, while all formats will be preserved:

"DANS-KNAW accepts and understands that the best standards and procedures for the storage, manipulation and access of digital materials are likely to evolve over time and that it will avail itself of the latest technologies in cooperation with ELSEVIER." (DANS contract)

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:

> The repository core infrastructural software is developed and maintained in-house, by a dedicated software development team, as described in R5.

> The repository uses metadata standards including Dublin Core and Google Science Datasets, for dataset metadata markup. Reference standards are an ongoing area for review and improvement, and we aim to add support for experimental and subject-specific metadata standards in future.

> The repository software is currently closed source and proprietary so documentation is not publicly available; however RESTful APIs are maintained which provide access to all public repository functions – these are publicly documented here.

> Funding has been allocated to support future infrastructure development in line with our needs.

> Mendeley Data is hosted, and research data stored, on Amazon S3, which provides a service across high-performance connections. Please refer to http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html for more information.

> Amazon S3 (Simple Storage Service) servers ensure a reliable and stable service at all times:

> • Backed with the Amazon S3 Service Level Agreement

- Designed to provide 99.999999999% durability and 99.99% availability of objects over a given year

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:

Mendeley Data takes a proactive approach to security of research data and user data.

Regular penetration testing is carried out to ensure service is secure against attack. All previous penetration tests have failed to breach the service; recommendations issuing from tests have been implemented.

Amazons AWS security is detailed here. Data is stored in a MySQL database hosted in an Amazon Relational Database Service (Amazon RDS) which has the ability to encrypt data at rest and/or in transit using SSL. Please refer to information here for further information.

Our data is stored with Amazon S3 rather than local Elsevier Servers, therefore our data falls under Amazon business continuity and disaster recovery arrangements.

Amazon Cloud Services SLA availability is over 99.9% over a given year and comes with cross-region replication. The data centres are within the EU, specifically Ireland.

Please refer to https://aws.amazon.com/ec2/sla/, https://aws.amazon.com/rds/sla/ and https://aws.amazon.com/s3/sla/ for further information.

Our AWS usage has been configured to use multi availability zones. AWS business continuity program can be found in https://d0.awsstatic.com/whitepapers/aws-security-whitepaper.pdf.

Additionally, our long-term archive partner DANS provides multiple backups and redundancy, from which datasets could be restored in the very unlikely event they could not be restored from Amazon S3.

Access to the repository functions, for ingest or access, are not currently dependent on Mendeley Data employee availability. Elsevier plans and maintains comprehensive security, disaster recovery, business continuity plans (including employee availability), and arrangements for Mendeley Data fall under these.

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:

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

Self-assessment statement:

Mendeley Data aims to interoperate with and contribute to the ecosystem of data repositories and other data services for example for discovery, persistence and access to data.

[Mendeley Data public (reviewed and valid) datasets are available here](#).

In terms of use by other projects, Mendeley Data datasets currently propagate across the following services:

- Indexed by [Data Search](#) – a cross-repository search engine developed by Elsevier

- Metadata records available in [DANS EASY archive](#)

- Metadata records available in [DataCite Explorer search index](#)

Open APIs serving dataset metadata and file download URLs enable indexing by other services.

In addition, our structured metadata, marked up with Dublin Core, and Google Science Datasets markup enable indexing. We are working with Dutch Techcentre for Life sciences (DTL) to make our data more FAIR (Findable, Accessible, Interoperable, Reusable), and able to be discovered via FAIR Dataports.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: