



Universiteit
Leiden

Research Data protocol Faculty of Humanities

Faculty: Humanities (Geesteswetenschappen)

Version number: 1.1.0

Status of this protocol: [~~Concept/Draft/Under Review/Approved/Other~~]

Contact person for questions about this protocol: [Myrte Vos](#)

Location of the latest version of this protocol: <https://zenodo.org/records/14938755>

Table of Contents

| | |
|---|----|
| Preamble to the data protocol | 3 |
| 1. Introduction | 3 |
| 2. To whom this data protocol applies | 3 |
| 3. Research data covered by the data protocol | 4 |
| Planning and preparing | 6 |
| 4. Collaboration with third parties | 6 |
| 5. (Re)using existing or third-party data | 7 |
| 6. Data Management Plans | 8 |
| Protecting and processing | 11 |
| 7. Secure management of data | 11 |
| Preserving and publishing | 13 |
| 8. Research data underlying a publication | 13 |
| 9. Research data not underlying a publication | 14 |
| 10. Preserving non-digital data | 15 |
| 11. Retention periods | 15 |
| Roles and responsibilities | 16 |
| 12. Management responsibilities | 16 |
| 13. Researcher responsibilities | 17 |
| 14. Where to find support | 18 |
| Appendix A - Further reading | 18 |
| 1. Relevant legislation, agreements, and guidelines | 18 |
| Appendix B - Glossary | 19 |

Preamble to the data protocol

1. Introduction

Good data management throughout the research lifecycle improves the transparency, integrity, reproducibility, and reach of scholarly research. It benefits the researcher, by supporting their workflow and mitigating risk of accidental data loss; their research community, by maximizing the utility and reusability of data, and protecting participants and interlocutors against data leaks; and finally, society at large, by ensuring long-term access to research data and making it available as a public good.

Individual researchers and institutions, but also research funders such as NWO and Horizon Europe are increasingly committed to making publicly funded research available through Open Access publishing, and archiving research data under the principle ‘as open as possible, as closed as necessary’. The Netherlands Code of Conduct for Research Integrity (2018) lays out researchers’ responsibilities with regards to data management, and institutions’ duties of care in providing the necessary information, infrastructure, storage capacity and support. Leiden University has formulated its policy in the Leiden University Data Management Regulations (RDM2021). This protocol is an elaboration of RDM2021 for the Faculty of Humanities.

1.1 Purpose of this data protocol

RDM2021 provides a framework that can be elaborated upon by faculties and Institutes. This protocol translates the clauses of RDM2021 into concrete implementation guidelines, specific to the type of research and practice in the faculty. The aim is to clarify the responsibilities of individual researchers and other stakeholders throughout the Institute, the faculty, and the university; and to encourage best practices in research data management.

1.2 Procedures for this data protocol

This data protocol takes effect as of March 1, 2025, and does not apply retroactively. The data protocol is reviewed when necessary and at least once every 2 years, by two or more research support staff members at the Faculty.

The data protocol is approved by the Faculty Board and Scientific Directors of the Faculty of Humanities.

2. To whom this data protocol applies

In line with RDM2021 (§3), this protocol applies to all employees and persons performing research under the auspices of Leiden University. This includes external PhD candidates and contract PhD candidates, visiting researchers, retired colleagues and any other guests or partners who carry out research at the Faculty of Humanities. This protocol does not apply to any work created by students for educational purposes; research conducted by BA and (non-Res)MA students falls under the formal responsibility of their supervisors.

3. Research data covered by the data protocol

3.1 What is research data?

Sabina Leonelli, philosopher and historian of science, defines ‘data’ as “a relational category applied to research outputs that are taken, at specific moments of inquiry, to provide evidence for knowledge claims of interest to the researchers involved. Data thus consist of a specific way of expressing and presenting information, which is produced and/or incorporated in research practices so as to be available as a source of evidence, and whose behaviour and scientific significance depend on the context in which it is used. In this view, data do not have truth-value in and of themselves, nor can they be seen as straightforward representations of given phenomena. Rather, data are essentially fungible objects, which are defined by their *portability* and their *prospective usefulness as evidence*.” (Leonelli, p. 811¹)

Within this view, whether entities function as data depends on the processes of inquiry in which they are used; and on the degree to which they can be disseminated, which in turn depends on their format and medium. The question ‘what is research data?’ can therefore only be answered with reference to *concrete research situations*, in which researchers make *specific decisions* about what can be used as evidence for which claims.

3.2 The scope of research data under this protocol

With the above in mind, this protocol scopes over all materials and information, digital and physical, that is collected or provides the basis for analysis in academic research, and that is necessary to substantiate and validate the outcomes of that research. This includes data in all stages of the research lifecycle - raw, processed, and analysed² - and the methods by which they are transformed from one stage to the next. It also includes software that is developed during the research – to support data collection, processing, and analysis, or as research output in its own right - and supporting documentation such as stimuli, codebooks, readme files, etc.

This scope is deliberately defined as broadly as possible, to be inclusive of the full range of epistemologies, methodologies, and research outputs found in the Humanities. It is primarily informed by what is needed to meet the twin responsibilities of the researcher to 1) corroborate claims with evidence, and 2) to do so honestly, scrupulously and transparently³. Secondly, it is aligned – where possible and reasonable – with the aspirations of the Open Science movement, striving to manage data such that it is accessible and interpretable to the widest possible audience.

Within the Faculty of Humanities, research data can include but are not limited to:

- Texts recorded, transcribed, or extracted from archives that are not (easily) accessible
- Video and audio recordings, photographs, and scans
- Surveys, forms, and questionnaires
- Physiological recordings (e.g. EEG, fMRI, heart rate, pupil dilation, motion sensor data)
- Behavioural measurements (e.g. eye gaze tracking, reaction time, error rates)
- Observational and analytical lab or field notes
- Annotations, trans- and descriptions, and translations of text, video, and audio files
- Selections from and adaptations of existing datasets (that are difficult to reproduce)

¹ Leonelli S. (2015) What Counts as Scientific Data? A Relational Framework. *Philosophy of Science*, 82(5): 810-821. <https://doi.org/10.1086/684083>

² See the Glossary (Appendix B) for the distinctions between these three ‘stages’ of research data.

³ See the Netherlands Code of Conduct for Research Integrity (2018)

- Corpus and database queries
- Experimental stimuli
- Computational models, algorithms, parameter settings, and simulations
- Digital data scraped from websites, social media apps, messaging apps etc.
- Software developed to collect, create, process, analyse, and present data
- Documentation of visual, musical, narrative and performance art, including interactive installations and social experiences facilitated by artists; and of artistic processes, including techniques, stages, and contexts of artistic creation, as well as materials and tools used to create artwork.

Research data management as outlined in this protocol mostly concerns digital data, but can include digitizing or digitally documenting research materials such as:

- Paper field diaries and lab notes, questionnaires, and correspondence
- Paper consent forms
- Hardware/setups used in laboratory experiments (including physical stimuli used in experiments such as toys and puppets), in art or museum exhibits, or during artistic performances
- Objects, documents and artefacts collected during fieldwork

Some physical materials or artefacts cannot be digitized, and so are at particular risk of loss or damage. Institutes may have physical storage and archiving capacity available; if not, consider including storage equipment (e.g. a fireproof safe) in research project (budget) planning.

3.3 Data outside the scope of this protocol

Researchers' field diaries and personal notes are not considered research data by default: in certain disciplines, such as anthropology and artistic research, they are fundamental to the research process and can constitute an important form of evidence, but they occupy a grey area between research data, academic writing, and private reflection. It is left up to the researcher to adjudicate which parts of her notes fall within what category.

Bibliographies and reference lists associated with academic publications resulting from a research project are not themselves considered research data within that project, unless they are the subject of scholarly inquiry.

Proof of informed consent, audit trails, statements of approval from the ethics committee, preregistration reports, grant proposal forms and other forms of administrative project documentation are not considered research data unless they play that role in the context of the research, or if the research data cannot be fully understood without them.

Planning and preparing

4. Collaboration with third parties

4.1 Knowledge security

A collaboration can be an interesting opportunity, but also bring risks: misuse or theft of knowledge, ethical issues associated with the application of research results, or unwanted interference that affects academic freedom. When the risks are substantial and/or complex, please reach out to the **Faculty Security Officer**, Leiden University's knowledge security helpdesk, and/or to the National Contact Point for Knowledge Security.

4.2 Legal agreements

When collaborating with third parties (be it other universities, public or government institutions, private companies, etc.), clear agreements need to be made on how research data will be collected, processed, accessed, used, and stored, as well as on intellectual property rights, such as copyrights and terms of use. When the collaboration involves personal data, it is very likely you will need a Data Sharing or Joint Controller Agreement: **contact the Faculty Privacy Officer as early as possible.**

When collecting personal data *through* a third party (i.e., where you are the data owner, and they are the data processor), you may need to draw up a data processing agreement (DPA) to comply with the GDPR. Online data collection tools and platforms for which the University has a license (e.g., Qualtrics) do not require a DPA.

Note that the researcher rarely has the mandate to sign legal agreements themselves. **Contact the Information Manager.**

4.3 Indigenous data sovereignty

Researchers at Leiden are obliged to conduct their research involving people in accordance with the principles of the GDPR, and with the rights of data subjects. This already gives research participants a relatively high degree of control over their personal data. However, some forms of research consider participants the (co-)owners of data gathered about them, or as co-creators of the research methodology and published output; or involve communities that assert authority to control data about them (see e.g., the CARE principles of Indigenous data governance). Such participant groups and communities may have a history of being exploited, mischaracterized, and marginalized by scientific research. In such cases, the researcher usually holds the research data 'in trust', and needs to not only proactively negotiate agreements around data access, licensing, publication, preservation, and authorship, but also center collaborators' needs and perspectives with regards to their data (e.g. in the choice of metadata terms or codes used during data analysis).

RESOURCES

Knowledge Security

1st point of contact:
Security Officer

2nd point of contact:
[Leiden Univ helpdesk](#)
[National Contact Point](#)

Legal agreements

Personal data?
Privacy Officer

Data use/sharing/transfer?
Information Manager

External partners?
[Luris](#) (knowledge
exchange)

Online data collection?
ICT/Software consultant
LUCL Lab Manager
Centre for Digital
Humanities
[SOLO lab support \(FSW\)](#)

Ind. Data Sovereignty

[GDPR Principles](#)
[Data Subject rights](#)
[CARE Principles](#)

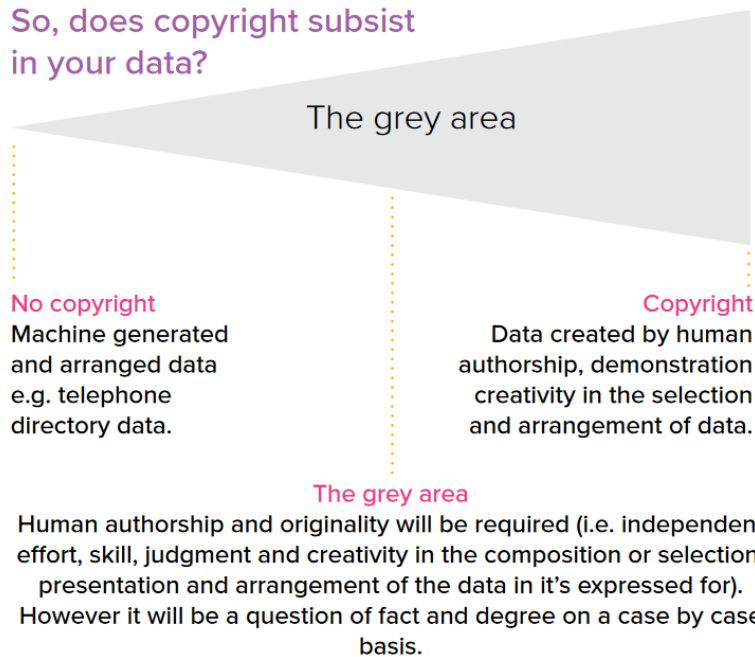
Digital tools for IDS:
[Provenance labels](#)

Further reading:
[State of Open Data \(ch.21\)](#)
[IDS \(2016 book\)](#)

5. (Re)using existing or third-party data

Third party research data is any data that has been created by other researchers or by external agents, such as census data created by Statistics Netherlands. This data may be sourced directly from the creator, or from secondary sources such as articles, books, corpus databases and repositories. Third party data may be protected by intellectual property rights such as licenses, copyright, patents and trademarks. If researchers reuse third party data, they should be aware of any restrictions.

So, does copyright subsist in your data?



(Figure from the ARDC Research Data Rights Management Guide)

Instead of an owner, research data usually has a 'rights holder' who determines its license and who may need to be asked for permission to reuse and publish it. Leiden University is generally the 'rights holder' of research data created under its employ (though this may depend on the specific contractual agreement with the researcher). In practice, researchers often set the license on their data and make decisions on how openly it is published.

Two exceptions to the copyright law are particularly relevant to researchers using digital research methods:

- You may make a private copy of a copyrighted work for research purposes, and
- You may reproduce copyrighted material that you have legal access to, for the purpose of **text and data mining**. You must store reproductions with appropriate security measures, and you may preserve them for academic research purposes (e.g., verification and replication of results).

RESOURCES

Intellectual property

For guidance on copyright law in Dutch higher education, see auteursrechten.nl, and [this guide](#) on the legal status of raw research data.

For specific advice on copyright and IPR, both as it pertains to your own research output and any existing data you want to (re)use, contact the [Copyright Information Office](#). [\[email\]](#)

Leiden University
[Employer Copyright Regulations](#)

Further reading

Australia Research Data Commons' [guide on data rights management](#).

Authorship and contributor roles

[CRediT](#)
[COPE statement on AI](#)

Scientific Integrity

NL Code of Conduct for Research Integrity
[\[NL\]](#)[\[EN\]](#)

[Complaints procedure and support](#) at Leiden University
[National Advisory Council for Research Integrity](#)

6. Data Management Plans

A data management plan (DMP) is a document that is developed at the start of a research project, which outlines all aspects of managing data both during and after research. A DMP should clearly describe all decisions and measures taken to guarantee responsible handling of research data and, if applicable, long-term availability of the research data.

A good DMP offers many benefits: it can structure conversations and aid decision-making about research workflows, especially in project teams and consortia, which saves time and mitigates the risk of costly misunderstandings and mistakes. It helps the researcher and their community (both in- and outside academia) to get the most out of valuable research data. Finally, it is a useful instrument in supporting research integrity.

Per RDM2021, **all** researchers at Leiden University are expected to create a data management plan at the beginning of the research process. **In practice, the Faculty of Humanities applies this policy as follows:**

- PhD candidates and grant-funded researchers are **required** to write a DMP for their project.
- Research projects which rely on well-documented source material, and do not generate new research data, are **exempt**.
- Researchers who are conducting research which is not externally funded, not intended for publication, and/or not linked to a specific project are **encouraged** to plan and document their data management strategy in whatever way best supports their work.

6.1 Templates that can be used for a DMP

Most large funding bodies have their own DMP template; some, like NWO, will also accept an institutional template. Leiden University has its own DMP template, but its use is not required. **For the time being, the template that is recommended by the Faculty Data Steward is the DMP template from NWO.** Researchers need to verify which template is accepted by the funder of their research.

6.2 Procedures for writing, consultation and approval of the DMP

The researcher is responsible for writing the DMP. Their collaborators and/or supervisors, if applicable, must be involved in (or at least kept informed of) the writing process.

DMPs for PhD research should be approved and signed by the project's primary supervisor. Data Stewards or librarians can be

RESOURCES

DMP templates

NWO [\[EN\]](#)[\[NL\]](#) – *preferred*

Leiden University
[\[template\]](#) [\[website\]](#)

European Research
Council [\[doc\]](#) [\[pdf\]](#)

DMP writing guides

ERC [guide for grantees](#)

Science Europe [guide](#)

[Open Science Guide](#) for
Early-Career Researchers

Making qualitative data
reusable: [a short guide](#)

FAIR data sharing in the
Humanities: [ALLEA report](#)

FAIR data in SSH [\[video\]](#)

[RDM for historians](#)

[RDM fieldwork checklist](#)

Support

1st point of contact:
[Data Steward](#)

2nd point of contact:
[Centre for Digital
Scholarship](#)

Join a workshop!
[Sign up here](#)

Further reading

[Philosophy of Open
Science](#) (Leonelli, 2023)

consulted during writing, but their approval is not required (with the exception of PhD candidates at LUCL, for whom it is part of their mandatory data management course). For grant-funded researchers, review and approval of their DMP by a local Data Steward (or similar research support officer) is commonly required by the funder. Note that the approval of the DMP will not set its text in stone: **the DMP should be a living document that adapts and refines as the research project progresses.**

In the case of research teams and consortia, it should be clearly agreed which research partner will be in charge of the overarching project DMP. Depending on the project, it may be practical to devolve certain aspects of data management planning down to the level of partner organizations or individual researchers, but this division of labour should be clearly outlined in the top-level DMP.

6.3 Procedures for registration, storage and availability of the DMP

For the time being, only PhD candidates and researchers with grant-funded projects are *required* to write a DMP. Concomitantly, only those two groups of researchers are required to submit, store, and ensure the availability of their DMP for at least the duration of their project.

- PhD candidates are required to submit their signed DMP, as well as later updated versions, to their Institute's Graduate Coordinator. The most recent copy of the DMP is kept on file with the candidate's graduate dossier in the Converis-GSM system. At the end of the PhD project, the final version of the DMP is archived together with the dissertation and the associated research data.
- Grant-funded researchers are required to submit their DMP to their funder (usually within a few months of the awarding of the grant). Updated versions of the DMP are clearly named as such and kept on file with their project administration. If required, the final version of the DMP is submitted to the funder as part of the final report.

RDM2021 stipulates a minimum 20-year retention period for DMPs. Appropriate digital infrastructure (e.g., software tooling, a database) to support the writing, review, and long-term storage of DMPs is currently under development. So long as this infrastructure is not available, the Faculty *encourages* researchers to write a DMP for their research workflow and store it alongside other vital project administration documents, but does not (yet) *require* that they submit it to their Institute or to the Faculty for internal storage.

6.4 Relation of DMP to other relevant procedures and documents

Data management planning, particularly for research that will involve human subjects and/or processing personal data, is closely entwined with data privacy and security considerations - which in turn form an important part of the Ethics Committee's review criteria. Researchers planning to apply to the Ethics Committee for review and approval of their project, are advised to write a DMP *first*. Projects involving personal data will need vetting, and potentially also tailored advice, from the Faculty Privacy Officer before the Ethics Committee can review them; this process goes smoother and faster when the Privacy Officer has access to a DMP.

Ethics review

The Faculties of Humanities and Archaeology share an Ethics Committee. Researchers who are planning to work with human

RESOURCES

Ethics review

Ethics Committee [website](#)
Checklist: [should I apply?](#)

1st point of contact:
[Secretary of the EC](#)

Further reading

[Data Privacy Handbook](#)
[Ethics Guidelines for SSH](#)
[Data Ethics Decision Aid](#)
[Ethics for ethnography](#)
[Social media research](#)
[Ethics in linguistics](#)
Deceased subjects [1][2]
Normen voor historici [NL]

Indigenous research:
[Guidelines for IR \(2012\)](#)
[Decolonizing IR \[1\]\[2\]\[3\]](#)

subjects or human remains; collect, process or analyse living people’s personal data; and/or cultural heritage artefacts and materials, are required to submit their planned research to the Committee for review. Review is optional for Research Master students, but recommended if the intent is for their research to be published, and/or if they conduct their research using facilities shared with the Faculty of Social Sciences (where review is mandatory).

Researchers present their research for review through the application form (downloadable through the website), plus any relevant supporting documentation, such as their DMP, an information and consent form, etc.

Privacy scan

A privacy scan is a list of questions used to determine the level of risk of privacy infringement and data leaks in a research project. A Privacy Officer will ask these questions in a consult and suggest mitigation measures, and advise the Ethics Committee during their review of the project. If the risks are deemed high, it may be necessary to carry out a Data Protection Impact Assessment (DPIA). ‘High risk’, in this context, means something else than ‘just’ confidential or sensitive personal data, or personal data collected under high-risk conditions (e.g. a warzone, a domestic abuse shelter). A DPIA is called for when you plan to, for example, process large volumes of ‘special category’ personal data (ethnicity, religious beliefs, sexual orientation etc.), evaluate people and/or make decisions about them based on automated profiling, or do large-scale monitoring of people in a public space. A DPIA is also recommended if researchers are working with new methodologies or technologies whose impact on the privacy of data subjects is not yet clear.



Figure from the UU [Data Privacy Handbook](#) [image link]

Note that guest researchers and external and contract PhD candidates are not contractual employees of Leiden University; therefore, the University is not formally, legally responsible for these researchers’ compliance with the GDPR. However, the Faculty of Humanities acknowledges its duty of care to *all* affiliated researchers, and encourages external PhD candidates to reach out for support and advice.

6.5 Updating the DMP

The DMP is a living document that is kept up to date during research. The researcher is expected to review the DMP at least annually and update it if applicable.

Protecting and processing

7. Secure management of data

7.1 Data Storage

The choice of storage location should depend, first and foremost, on the sensitivity and protection needs of the data. A *confidentiality, integrity and availability* classification (in Dutch: BIV-classificatie) of data storage options at Leiden University is under development.

Additional safety measures need to be taken for the storage/handling of confidential or sensitive (personal) data, including data that constitute a security risk for the University or other parties. **As of the current version of this protocol, Leiden University does not offer appropriately secure storage for high-risk sensitive data.** Contact support staff to discuss alternative solutions.

Leiden storage

Leiden's internal storage provisions (P:drive and J:drive for single- and multi-user storage, respectively) are currently being slowly phased out, in favour of institutional Microsoft 365 applications (OneDrive, Teams, Sharepoint) and SURF applications (SURFdrive, Research Drive).

Note that the 'grace period' for departing researchers' UCLN accounts is 60 days (two months)! After this period, the account and its associated stored data are destroyed. A departing researcher and their supervisor have a joint responsibility to negotiate and transfer stewardship of any (copies of) research data that was collected or generated during the researcher's employment term.

External PhD candidates

By default, guest researchers and external PhD candidates have limited access to ICT resources at Leiden University. Nonetheless, the security needs of their data may require a managed device, encrypted storage, specialized software, or other measures. Contact support staff to discuss available options.

7.2 Data transport and transfer

Researchers should transfer data directly, or as quickly as possible, to their recommended/approved data storage solutions. When transporting or transferring sensitive data, the device or transfer protocol must use encryption.

University-managed hardware and software should be used wherever possible: e.g., when interviewing participants remotely, use Teams instead of Zoom; and when sharing files with other researchers, use OneDrive or SURFdrive rather than Dropbox.

RESOURCES

Research ICT

[Research Support Portal](#)
[Data Storage Finder](#)

Data collection

[SOLO research wiki](#)
[LUCL lab support](#)
[LUCDH](#)

Safe browsing

[eduVPN](#)
[Tor browser](#)

Encryption

7zip, [VeraCrypt](#)

Data transfer

[SURFfilesender](#)
[Wormhole](#)
[Keka](#)

De-identifying data

[Data Privacy Handbook](#)
[Infographic](#)
[Software tools](#)

Support

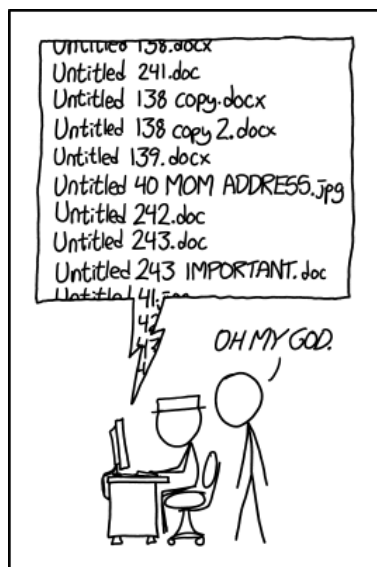
1st point of contact:
Software Manager
Data Steward

2nd point of contact:
Information Manager
Research ICT Coordinator
Security Officer

3^d point of contact:
[Helpdesk](#) > Research
Support

7.3 Data documentation

Stored datasets need to be well documented. This includes a brief description of the dataset (metadata properties), an overview describing individual files, and (where applicable) codebooks describing data elements used and descriptions of code syntax. Clear and transparent documentation helps save time both during a project (the code written as a stroke of genius at 2 AM may, when reconsidered a month later, no longer be intelligible to its author, let alone anybody else), and at the end of the project, when data is prepared for publishing and archiving.



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

[xkcd #1459](#) (Randall Munroe)

7.4 Version Management

Proper version management should be practiced such that the original (raw, unmodified) version of the data can always be identified; versions of the data that constitute the basis for specific publications can be identified; and errors when working with the data can be mitigated so that a minimum (e.g., one day's worth) of work is lost.

Exceptions (for instance because of the size of a dataset or security concerns) from this general guidance should be motivated in a DMP.

7.5 Data formats

Data should be archived using file formats that offer the best long-term guarantees for usability, accessibility and sustainability. As a general guideline, such preferred formats:

- are frequently used;
- have open specifications;
- are independent of specific software, developers or vendors.

For example, use .odt, PDF/A or .txt formats for text documents rather than .doc(x) or .rtf; .csv for tabular data instead of .xls(x), etcera. Or convert to sustainable formats at the end of the project. In practice, it is not always possible to use formats which satisfy all of these criteria. It may also sometimes be desirable to archive data in 'non-preferred formats' instead of (or ideally in addition to) preferred formats: i.e., formats that are widely used in your field, and which will be moderately to reasonably usable, accessible and robust in the foreseeable future.

RESOURCES

Metadata

[Metadata intro](#)

[Metadata catalogue](#)

[How to write a Readme](#)

[DublinCore generator](#)

[Data Curator app](#)

File management

[File and folder naming](#)

[Sustainable file formats](#)

[Version management](#)

Documenting your work

[Reproducible data analysis](#)
(tutorial for language data)

[nodegoat](#) is a web-based data management, network analysis and visualisation app for the Humanities. Leiden University has a license.

Custom data engineering

1st point of contact:

[Research software engineer](#)

2nd point of contact:

[Centre for Digital Scholarship](#)
[Centre for Digital Humanities](#)

Preserving and publishing

8. Research data underlying a publication

8.1 Archiving data underlying publications

Per RDM2021: “Digital research data that form the basis for a scientific publication are registered at the time of publication and managed according to the FAIR principles.” ‘Registering’ research data is taken to mean that data should be archived in a trusted, and preferably certified, research data repository. A publication with an associated dataset should include a persistent identifier to the archived dataset, and vice versa.

The first and foremost purpose of archiving data is to support the principles of research integrity, by facilitating the verification of researchers’ claims. By archiving data, researchers also contribute to a more open, sustainable, and equitable research landscape, by making their data easier to access (including by non-academic audiences), cite and reuse.

There may be legal, ethical, commercial, political, or cultural reasons for not archiving data openly/publicly, including but not limited to personal data protection, intellectual property rights, knowledge security risks, and (indigenous) communities’ authority to control data from or about them. When it is not possible to archive data with open/public access, it should be archived with restricted or closed access, so that the dataset’s existence can be discovered by others through its metadata, and access requested and granted as and when appropriate. When data is exceptionally sensitive, contact research support staff to discuss options.

8.2 What to include in an archived dataset

Not *all* data and digital material associated with a publication merits long-term archiving. Criteria that researchers can use to curate their dataset for preservation include:

- Is it really research data? (e.g., conference abstracts or presentation slides *about* the data)
- Are the data already archived as part of another project, or with a publication?
- Are the data directly relevant to the publication(s)?
- Are the data unique (impossible to recreate)?
- Are the data valuable? (E.g., in terms of transparency, reuse, quality, originality, size, possibility to combine them with other data, production costs or innovative nature?)
- Are there any obligations for long-term storage?
- Are the data subject to any publishing restrictions?
- If long-term preservation is chosen for the data, is the chosen infrastructure adapted to their characteristics (estimated cost, size, desired accessibility during preservation, etc.)?

RESOURCES

Archiving data

[DataverseNL](#) – *preferred*
Dataverse User Manual for
Humanities researchers

[Research Catalogue](#)
(ACPA)

[Repository finder](#)
[DANS Data Station SSH](#)
[The Language Archive](#)
[Special Collections UBL](#)

Publishing data

[Research Data Journal](#)
[Zenodo](#)
[Open Science Framework](#)

Find a usage license

[Creative Commons](#)
[Software licenses](#)

Non-digital data

[OpenAIRE guide](#)

Support

1st point of contact:
Data Steward
Dataverse Curator

2nd point of contact:
[Centre for Digital
Scholarship
Documentary Information
and Archiving \(DIA\)](#)

Further reading

[Open Science Guide](#) for
Early-Career Researchers
[Challenges](#) of qualitative
data sharing in SSH

8.3 Preferred / certified repositories

Data should be deposited in a trusted repository (according to [TRUST principles](#), ideally with CoreTrustSeal-certification). Researchers are free to choose a repository that best suits their project (for example, the Endangered Languages Archive for documentation of a nearly-dormant language, or a municipal archive for street photography or urban field recordings), but as a default, they are encouraged to use Dataverse repository of the Institute. Within the Leiden instance of DataverseNL, the Faculty of Humanities maintains DataverseNL repositories for every Institute (except NIMAR, which is categorized under LIAS), and one for the Centre for Digital Humanities. Dataverse supports access restriction at the file level, and the creation of anonymized peer review links.

DANS also maintains a Data Station for Social Sciences and Humanities datasets, to which any researcher regardless of institutional affiliation may submit their data (for free up to 50GB). For the time being, DANS assists in the curation of such datasets. Note that if the dataset you wish to store at DANS contains personal data, the Faculty's Information Manager will need to sign a [Data Processing Agreement](#).

Note: under GDPR, personal data must be stored within the European Economic Area. Check whether your repository is hosted or maintains a server in the EEA (for example, the Open Science Framework is administered by an organization based in the USA, but allows users to select a German storage location for this reason).

8.4 Choosing a usage license

Published research data should come with clear and accessible access conditions, and a data usage license. It is recommended to choose a usage license that makes data available to the widest possible audience, and allows the widest possible range of uses.

9. Research data not underlying a publication

9.1 Archiving data not underlying publications

At the Faculty of Humanities, it is currently *encouraged*, but not *required* to archive data not underlying publications. Unpublished data may nevertheless be worth preserving. The same appraisal criteria that are listed in §8.2 apply; please contact the Faculty Data Steward, and/or the Dataverse Curator, to discuss whether your dataset is suitable for archiving.

9.2 Archiving data underlying BA and (non-Research) MA theses

As stated in §3, this protocol does not apply to students *unless* their research results in an academic publication (in which case the recommendations given in the previous sections apply). The minimal retention period of 10 years (see §11) likewise does not apply to this group, although it is advisable for data underlying a thesis to be retained, and made available to examiners on request, until after graduation.

However, many students do create or collect original data in the course of writing their thesis and may wish to preserve it. If the data's size and format suit it, data can be included in full as appendices to the thesis and so be archived along with the manuscript in the Leiden Student Repository. If this is not practical, the data may be appraised by the thesis supervisor and the curator of the relevant institutional Dataverse repository, to assess whether the data and its documentation is of high enough quality to be archived there. Alternatively, the student may archive their data in a project repository on e.g., the [Open Science Framework](#) and include the associated DOI in their manuscript.

In line with the Leiden Student Repository's archiving policy, BA theses (and by extension, the underlying data) are closed access by default; MA theses are open access by default, unless they are under embargo. The data underlying MA theses can be open, restricted or closed, irrespective of the access conditions of the manuscript. The author and thesis supervisor are jointly responsible for ensuring that archived data does not contain any personal data.

10. Preserving non-digital data

The storage and preservation of non-digital data – if any – is organized at the Institute level. Researchers who are unsure what to do with their non-digital data at the end of the research project, or their career, are encouraged to contact their Institute Manager and Faculty Data Steward to appraise the material and discuss options for safeguarding.

10.1 Archiving legacy data

Legacy data refers to information that is stored in outdated or obsolete systems, formats, or technologies that is often difficult to access. Its provenance and current ownership may be unknown. This protocol does not (yet) prescribe any particular way of managing legacy data, but scientific and support staff are encouraged to identify and appraise legacy data when they find it, and to notify the faculty Data Steward for further assistance.

Researchers leaving Leiden University (including emeriti) are urged **not** to abandon their legacy data at their Institute without a) explicit agreement with management staff about what should happen to it, and b) sufficiently describing the data such that whoever assumes responsibility for it can understand and manage it.

10.2 Archiving a personal collection

Researchers with an outstanding career scholarship record, and/or whose personal collection of non-digital data (e.g. notebooks, letters, sketches, rare books) can be considered to have (scientific-) historic value, may be able to donate their collection to the Special Collections department of the University Library (UBL). Even if the collection cannot be re-homed with the UBL, it may well be relevant to the acquisition mandate of another foundation, library, or archive. In addition, Digital Scholarship Librarians at UBL may be able to assist you in digitizing (parts of) your collection, or transforming your data into a searchable archive or database (as in [this example](#)).

11. Retention periods

As stipulated by RDM2021, the default *minimal* retention period for research data is 10 years after the publication related to the data appeared. The archiving period can be longer, depending on legal requirements, discipline-specific standards, and the intended purpose(s) of archiving the data. When no particular restrictions apply, a dataset will commonly be archived without a *maximum* retention period; however, if a maximum retention period applies, there is an obligation to destroy the data after the expiry date. Destruction must be done properly, in accordance with Archive law; please contact DIA (Documentary Information and Archiving) for support.

When the research data includes personal data, the GDPR principle of data minimization must be applied as soon as possible. That is, the data controller (in this case, the researcher) should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose. They should also retain the data only for as long as is necessary to fulfil that purpose.

Roles and responsibilities

12. Management responsibilities

12.1 Executive Board

The Executive Board establishes the policy frameworks for data management (at present: RDM2021) and evaluates these on a regular basis. The Executive Board provides adequate central facilities and support at central level to facilitate responsible data management. The Executive Board monitors compliance with the Data Management Regulations on a regular basis.

12.2 Faculty Board

The Faculty Board is responsible for:

- Providing the means and support for the elaboration, implementation, review, and regular evaluation of the Faculty data protocol. Ensuring that review and evaluation is scheduled at least once every two years, or whenever revisions are needed to remain compliant with any governing regulations and policies.
- Deciding on Faculty policy regarding matters that are underspecified by University policy.
- Delegate their responsibilities or give mandate where relevant to the research portfolio holder (currently, the Dean of the Faculty).
- Fostering recognition and awareness of responsible research data management within the Faculty, and for developing and disseminating its own vision on research data management to staff.

12.3 Scientific Director

The Scientific Director is accountable to the Dean of the Faculty, and is generally responsible for research data management within their Institute. Note: this responsibility may be delegated to the Research Director (if applicable) of the Institute.

The Scientific Director is responsible for:

- The elaboration and regular review and updating of the institutional appendix to the Faculty data protocol; and for the awareness and adoption of the data protocol within the Institute.
- Making final decisions (or delegating that mandate to appropriate scientific or research support staff) with regards to research data produced within the Institute, such as decisions regarding data access, auditing, destruction, etc. This includes mediating disputes about research data ownership and integrity.

12.4 Supervisors and managers

The supervisors and managers are accountable to the Scientific Director. They are responsible for:

- Developing, maintaining, and disseminating research data management procedures for their section or team (e.g., facility managers, collection managers, data managers, research software engineers, etc.).
- Being aware of where their team members store their files, and making sure that if necessary, access to and/or stewardship of their data is transferred to them (or another appropriate staff member) at the end of a research project or an employment term.

13. Researcher responsibilities

13.1 Principal Investigator / PhD supervisor

Where the role of Principal Investigator overlaps with the role of Researcher, please refer to §15.2.

Principal Investigators are responsible for:

- Ensuring at the beginning of the research project that there is clarity about data ownership and data management responsibilities, and that these are documented in any associated data management plan (DMP) and/or collaboration agreement.
- Creating and developing a DMP; updating the DMP as and when needed; and ensuring that all project collaborators are aware of and adhere to the most recent version of the DMP.
- Archiving the DMP, and overseeing the appropriate long-term storage or archiving of the research data and project documentation at the end of the research project.

PhD supervisors are responsible for:

- Ensuring that their PhD candidates have access to the necessary knowledge, skills, and resources to manage their research data appropriately.
- Supervising their PhD candidates' data management planning and ethics review applications; monitoring whether their work is carried out along the lines set out in their DMP and ethics review documentation; and ensuring any significant changes are reported to the ethics committee and reflected in updates to their DMP.

13.2 Researcher

Researchers are expected to adhere to RDM2021 and to the faculty and/or Institute data management protocol. Beyond that, they are responsible for:

- Familiarising themselves and complying with the relevant data management, record-keeping, open data, and retention requirements of research funders, sponsors, publishers, and other relevant external stakeholders, as well as the relevant legal, ethical, and regulatory frameworks governing the processing of personal data and sensitive research data.
- Making appropriate decisions and applying best practice in relation to the management of their research data and research related records. All research can benefit from research data management planning, and researchers are strongly encouraged to create a DMP before starting any research project.
- Making appropriate decisions relating to the retention, transference of ownership, or destruction of research data held in the university's storage environments if/ when they leave Leiden University. Responsibility for curatorial decisions thereafter (e.g. the deletion or migration of data) lies with the Principal Investigator or PhD supervisor (in the case of PhD researchers). Failing that, responsibility transfers up the management chain and terminates with the Scientific Director of the Institute.

14. Where to find support

For up to date information on available research support staff and services, please consult the [Research Support Portal](#).

14.1 Training and information

For training on research data management, please sign up for the Centre for Digital Scholarship's regular workshops. These also include support in writing a data management plan, and can be adapted to a specific Institute or research domain. The Faculty Data Steward is also available for data management consultations.

Appendix A - Further reading

1. Relevant legislation, agreements, and guidelines

A detailed list of the relevant legislations, both broadly applying to all scientific disciplines (elaborating on RDM2021, §2) and specific to the Faculty of Humanities, is given below:

Dutch and EU legislation:

- WMO (Wet Medisch-wetenschappelijk Onderzoek met mensen) [\[NL\]](#)
- GDPR (General Data Protection Regulation) [\[EU\]](#) [\[NL\]](#)
- Wet op het Hoger onderwijs en Wetenschappelijk onderzoek (WHW) [\[NL\]](#)
- Archiefwet [\[NL\]](#)
- Auteurswet [\[NL\]](#)
- Databankenwet [\[NL\]](#)
- Wet op de naburige rechten [\[NL\]](#)

(Inter)national guidelines

- 2018 Netherlands Code of Conduct for Research Integrity from NWO [\[EN\]](#)[\[NL\]](#)
- 2018 Code of Conduct of the National Ethics Council for Social and Behavioural Sciences [\[EN\]](#)
- 2021 European Commission guidelines for Ethics in Social Science and Humanities [\[EN\]](#)
- 2021 Science Europe Practical Guide to the International Alignment of Research Data Management [\[EN\]](#)
- The CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics) [\[EN\]](#)
- The FAIR Principles (Findable, Accessible, Interoperable and Reusable) [\[Article\]](#)
- The FORCE11 and RDA Joint Declaration of Data Citation Principles [\[EN\]](#)
- 2025 Dutch Research Council Policy on the use of Generative AI [\[NL\]](#)

University guidelines

- 2021 Research Data Management Regulations of Leiden University [\[EN\]](#)[\[NL\]](#)
- Leiden University policy on privacy and information security [\[NL\]](#)

Faculty-specific guidelines

- Regulation of the Ethics Committee of the Faculty of Humanities and the Faculty of Archaeology [\[EN\]](#)[\[NL\]](#)

Appendix B - Glossary

CARE Principles for Indigenous Data Governance

The CARE Principles (Collective Benefit, Authority to Control, Responsibility and Ethics) complement and challenge the FAIR principles for open science by encouraging sensitivity to power differentials and historical contexts, and centering benefit to and sovereignty of Indigenous people in the management of Indigenous research data.

Certified repository (Trusted digital repository)

A data repository is an archive for research data. A trusted digital repository is a digital archive whose mission is to store, manage and provide reliable, long-term access to digital resources and it has been certified by an official organisation. A well-known certification for data repositories is Core Trust Seal.

Copyrights

Raw research data is not protected by copyright. Copyright law only provides protection for works that display a certain creativity and originality. You can be the copyright holder in respect of certain data if the data, due to their form, qualify for protection (see [this report](#) from the Centre for Intellectual Property Law). Collaboration or consortium agreements, user rights, data user agreements or licenses are used to define what others are allowed to do with research data.

CoreTrust Seal

Core Trust Seal is a certification for repositories engaged in long-term preservation and sharing of research data. The certification evaluates repositories' technical infrastructure and standards, their organizational, financial, staffing and legal aspects as well as their workflows and risk management.

Creative Commons license

Creative Commons licenses are a standardized way to grant the public permission to use your creative work under copyright law. From the reuser's perspective, the presence of a Creative Commons license on a copyrighted work answers the question, *what can I do with this work?*

De-identification (ano- or pseudonymization)

The process of removing all or part of the identifying information, or separate the identifying information from the research data in order to make (indirect) identification more difficult.

- **Anonymisation:** the process by which personal data is altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party. The process must be irreversible.
- **Pseudonymisation:** a de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers, or pseudonyms.

DANS

Data Archiving and Networked Services, the Dutch national centre of expertise and repository for research data. DANS is a subsidiary of KNAW, the Dutch National Academy of Sciences. It curates and maintains 4 thematic Data Stations that any researcher can deposit datasets to, and manages the technical infrastructure for DataverseNL.

DataverseNL

DataverseNL is a research data repository co-provided by DANS and participating institutions. DANS manages the technical infrastructure, and the institutions using DataverseNL are responsible for granting rights to user accounts, managing and curating the deposited research data within DataverseNL. DataverseNL runs on open source software from the Dataverse Project, which was initially developed and still principally maintained at Harvard University.

Several faculties at Leiden University offer DataverseNL; the Faculty of Humanities offers subdataverses for each Institute (except NIMAR, which is categorized under Area Studies), plus the Centre for Digital Humanities. A user manual for submitting datasets to the Faculty Dataverse can be found at the Research Support Portal.

Data (raw / processed / analysed)

- **Raw data** are the original data that you have collected but have not yet processed or analysed. This can include third party data, i.e. data you did not collect yourself and are re-using.
- **Processed data** are the data that you have digitised, translated, transcribed, cleaned, validated, checked and/or anonymized.
- **Analysed data** are the models, graphs, tables, texts etc. created from the raw and the processed data, that are intended to aid in the discovery of useful information, the presentation of conclusions, and decision-making.

Data preservation

Data preservation refers to the series of managed activities necessary to ensure continued access to digital materials. These actions include but are not limited to the appropriate selection and appraisal of research data and research related records for preservation. In this protocol, data preservation is considered synonymous with data archiving.

Data publication

The publication of research data, either in a repository or in a journal dedicated to publishing data (or software). Data can be published with full or restricted access, immediately at the end of a project or after an embargo period, with a more or less permissive use license, etc.

Data storage

Data storage refers in this protocol to the storage of research data while the research project is ongoing.

Data Management Plan (DMP)

A formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation.

Data Protection Impact Assessment (DPIA)

Assessment procedure for identifying risks according to the GDPR's privacy principles and compliance requirements.

Embargo period

During an embargo period, only the description of the dataset is published, while the data themselves are closed. In most cases, a reasonable embargo period is 6 to 24 months.

FAIR principles

The '[FAIR Guiding Principles for scientific data management and stewardship](#)' (2016, *Scientific Data*) are a set of 15 principles aimed at improving the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets (see www.go-fair.org). The principles emphasize machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with minimal human intervention) because humans increasingly rely on computational support to deal with data.

FAIR is aligned with the Open Science movement, and thus often conflated with evolving standards of methodological transparency and research integrity. It is important to note that **FAIR is not equivalent to open**: there are many reasons why data may be restricted and only available under certain conditions to certain users, including machines. As long as the accessibility conditions are properly described, non-open data can be entirely FAIR.

Findability of data

To ensure the findability of data, researchers are encouraged to deposit their data in a repository that (1) grants a persistent identifier (for example a DOI) to the dataset; (2) that provides or allows for rich metadata corresponding to the dataset; and (3) that indexes this metadata such that it can be found via multiple search channels (e.g. within the repository's own catalogue, but also via general search engines on the web).

Accessibility of data

The accessibility of data is a function of (1) the ability of search tools to retrieve (meta)data using persistent identifiers; and (2) the existence of and adherence to procedures that outline who, and under what conditions, can access a dataset. Even in cases where a specific dataset has been deleted, destroyed, or is otherwise no longer available, the metadata for that dataset should remain accessible via the same channels, and under the same conditions initially used to access the dataset itself.

In practice, the accessibility of the dataset can be ensured mainly by choosing a trustworthy repository. The repository should ensure that:

- the access conditions of the data are clearly displayed;
- if an embargo needs to be applied, only the description of the dataset is published;
- standardized exchange protocols are used so that metadata are publicly accessible and harvestable by machines.

Interoperability of data

To ensure research data is interoperable with other datasets as well as common workflows for data analysis, storage, and processing, researchers are encouraged to use:

- Well-known and preferably open file and software formats;
- Relevant metadata, including community-standard schemas, controlled vocabularies, etc.

Reusability of data

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

To facilitate reusability of the data, the researchers can make sure that:

- Any data comes with a clear and accessible data usage licence;

- Files within the dataset are named according to clear and well-documented naming conventions (including versioning);
- The rich metadata meet research-domain relevant community standards, if any exist.

General Data Protection Regulation (GDPR)

European legislation governing the processing of personal data. In the Netherlands, these are elaborated as the 'Uitvoering Algemene Verordening Gegevensbescherming (AVG)'.

Metadata

Metadata are information that describes data. Metadata provide context and meaning to your research data or research materials and enable you and others to make sense of and reuse your data in the future. At a minimum, your documentation or metadata should clearly tell the story of how you created, gathered, and used your research data, and for what purpose. Depending on your field, or the specific aims of your research project (e.g. contributing your data to an archive, or a larger research consortium), there may be a pre-existing metadata standard you can or should adhere to.

[Rich metadata](#) corresponds to an accepted metadata standard in a machine-readable format; generally, a more elaborate and specific set of terms than the commonly accepted minimal standard offered by [Dublin Core](#) and [DataCite](#). In practice, the creation of such a rich metadata file consists of filling in a simple form that will then produce a metadata text (that can be saved as a readme file if necessary). Some repositories propose to automatically create rich metadata when registering a dataset. If not, the researcher can use a pre-existing rich metadata generator.

ORCID

All researchers affiliated with Leiden University are required to create an [ORCID](#), another example of a persistent identifier. However, whereas a DOI is used to find and identify discrete research outputs such as a publication or a dataset, an ORCID is a unique series of numbers assigned to an individual researcher. The ORCID is particularly useful for (1) finding and identifying an individual researcher; (2) linking that individual to the contributions they have made to the scholarly record; as well as (3) disambiguating the identities of several authors who may share the exact same name. ORCIDs are also increasingly accepted as login authentication for various research platforms, such as the Open Science Framework, OpenAIRE, Overleaf, etc.

Persistent Identifier (PID)

Persistent identifiers can be defined as long-lasting references to a digital resource. They reliably point to and unambiguously and uniquely identify a digital entity.

Examples of PIDs are:

- DOI for articles, datasets, etc.;
- ORCID for authors;
- Grant-ID for grants;
- ROR for institutions;
- SWHID for software

Personally identifiable data / sensitive data

Personal data, as defined by GDPR, involves any information that can be traced back to a specific person. This includes not only names, addresses and places of residence, but also bank account numbers, telephone numbers and postal codes with house numbers. Data such as IP addresses and browser settings are also considered personal data in certain instances. This is the case if you can identify someone based on this data (whether or not in combination with other data). Sensitive data

such as a person's race, religion or health are called special category data. They have been awarded extra protection by the legislator.

Replicability

Replicability refers to obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

Reproducibility

Reproducibility refers to obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

Research integrity

As defined by the Netherlands Code of Conduct for Research Integrity, research integrity is based on five guiding principles: Honesty, Scrupulousness, Transparency, Independence and Responsibility.