



Faculty of Humanities Research Data Management Policy

July 2020

About this document

This document sets out the policy on the management of research data (hereinafter data) at the Faculty of Humanities of Utrecht University. This policy outlines the criteria and guidelines for adequate data management and specifies the various roles and responsibilities of the parties concerned at the faculty.

The document is divided into three parts:

- I. A description of the aim and scope of the data management policy as well as the policy frameworks within which the policy applies.
- II. A description of the definitions, principles, responsibilities and guidelines for the policy.
- III. A description of the current support for data management at the faculty and the university.

I - Aim, scope and frameworks

Aim of this policy

The aim of this policy is to ensure careful data management within the faculty during and after the completion of research projects. The faculty regulates data management according to the FAIR principles:

- **Findable:** data and metadata are findable, preferably in a digital system.
- **Accessible:** data and metadata are stored and maintained in such a manner that they can be accessed and downloaded easily. In addition, there are clear provisions on accessibility and use.
- **Interoperable:** data and metadata are available in common formats and conform to regular standards so that the exchange and the combining of such data with other data or metadata is optimally facilitated.
- **Reusable:** data are described in a manner that facilitates research reproduction and/or the combining of the data with other data.¹

The policy focuses on digital data. The faculty aims to store research data in a digital format, as much as is possible. Particular focus is placed on data containing personal data or data to which intellectual property provisions apply.

Scope of this policy

The policy applies to everyone who is affiliated with the Faculty of Humanities at Utrecht University and conducts research, including internal and external PhD candidates, but explicitly excluding students.

Rules, policy frameworks and conditions

This policy represents the further implementation of a number of laws, regulations and policy frameworks, primarily the Utrecht University Policy Framework for Research Data and the Netherlands Code of Conduct for Research Integrity, in which the General Data Protection Regulation (GDPR) has been taken into account. Furthermore, the policy is in line with the wishes and requirements of research grant providers in the area of data management, in particular the

¹ Reproduction means: performing the analysis again using the same dataset. Replication means: conducting the research again, including the data collection.

Netherlands Organisation for Scientific Research (NWO) and the European Research Council (ERC), and the principles of Open Science.

II - Definitions, responsibilities, principles and guidelines

Definitions

Research project

In this document a research project means: all the activities conducted by researchers aimed at arriving at a scientific analysis or opinions, irrespective of whether they are reflected in or are geared towards publications (or other forms of output).

Research data

Research data are all the data or procedures (such as algorithms) used for a research project that form the basis for the results or lead to results.

Research data may take many forms including but not limited to: observations, documents, literature references, archaeological material, books, transcribed interviews, source codes, software scripts, parameter settings, theoretical models, simulations and measurements. All the material needed to reproduce research results is regarded as research data and requires adequate data management.

It is common practice to differentiate between unprocessed (primary/raw/rough) data, the procedures that cleanse, anonymise and prepare the data, and the final secondary 'cleansed' data.

Personal data

The GDPR defines personal data as: any information relating to an identified or identifiable natural person who is still alive. This means information that either relates directly to a person or can be traced back to that person.² Personal data can also be provided to third parties, for example, during interviews or in open questions for surveys.

Research projects in which interviews or surveys are used often involve personal data, i.e. the contact details of the research population. Even though personal data usually are irrelevant for the research, they do form part of the total research data and partially safeguard the integrity of the research.

Data package

A coherent set of data and procedures that has been stored unchanged for a longer period of time (archived). A data package contains a description, also referred to as metadata, which clearly shows the content of the package. Based on the data package and the accompanying metadata, it must be possible to validate the claims that are based on the data concerned.

A data package can consist of different data, such as multiple tables, scripts for statistical processing, stimulus material, audio files containing interviews and the interview transcripts in text files.

Archiving (of data packages)

This refers to the storage of data in an unchanged form for a longer period of time. In many cases data are packaged together with a code book, a lab journal (or archive journal) and a description of the data, the metadata. The set of data and the accompanying description are referred to as a data package. Data packages are archived at the time at which a scientific publication is based on the data package or at the end of a research project.

Publication (of digital data)

² <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/mag-u-persoonsgegevens-verwerken#hoe-weet-u-of-u-persoonsgegevens-mag-verwerken-6310>

If the data are archived as a data package, the package can subsequently be published. This means that the data package will be findable. In practice, the publication of a data package means that the metadata are published (in a data catalogue, for example).

Although publication makes a data package findable, this does not mean that the data are automatically accessible: a data package can be published as a 'Closed Access Package' (see also p. 6). A persistent identifier, a reference that enables the clear referencing and easy localisation of digital data, is assigned upon publication of the data package.

Responsibilities

A general principle is that every staff member is expected to manage relevant data as a good steward. In addition, Utrecht University specifically distinguishes between the following actors, each bearing their own responsibilities for data management.

1. The data creators: researchers and PhD candidates and support staff, who have primary responsibility for looking after the research data. In the case of students, the responsibility lies with their supervisor(s). Supervisor(s) ensure that students are aware of the guidelines on responsible data management. In the case of support staff, the responsibility lies with the researcher ultimately responsible for the particular project.
2. Directors of education and research directors are responsible for ensuring that the researchers, research leaders, students and their supervisors in the department abide by the university policy framework and any supplementary faculty regulations. They may also establish further guidelines as needed.
3. Deans draw up further faculty guidelines pursuant to the university policy framework and are responsible for ensuring that the directors of education and research directors implement both the university policy framework and the faculty guidelines.
4. The Executive Board draws up, ensures monitoring and facilitates the implementation of the general university framework, for example, through the provision of infrastructure, information and other forms of support.

Principles

Data storage and archiving

The management of research data often relates to forms of storage. A distinction is made between the storage of data used in ongoing research projects and archival storage (see p. 6).

The storage of data used in ongoing research projects relates to data and scripts that have been created, adjusted and possibly deleted again during the course of the research project. In archival storage the data are stored unchanged for a predetermined period of time.

The principle underlying the faculty policy is that archival storage must be used as soon as the data concerned are used to substantiate a scientific claim, for instance, for a scientific publication.

Rights

- Utrecht University has the authority to make decisions on data packages stored under the management of Utrecht University on its own infrastructure to the extent that these decisions do not conflict with earlier agreements made with the researcher(s) or between the researcher(s) and other rightholders. Examples are: transferring the data package to another data archive (such as DANS Narcis) or the migration of file formats.
- Exclusive rights for the reuse or publication of research data will not be transferred to commercial publishers and their representatives without the owners themselves retaining the right to make available the research data for reuse through open access, unless this originally was a precondition for the provision of funds.

- Any researcher who is responsible for a research project must draw up a data management plan (DMP) in any event prior to publication, but preferably before the start of the research project. The DMP should at least describe what data will be collected as part of the project, where the data will be stored and who has access to the data during the research project. It also states where (a version of) the data is archived as a data package, how the data package can be found, who can access it and under what conditions.
- If personal data are used in a research project, the DMP should explicitly describe what additional measures have been taken to ensure the lawful and careful handling of personal data. This might include a description of the legal basis for collecting the data (e.g. informed consent), the anonymisation method, and processing upon completion of the project (e.g. the destruction of contact details). Furthermore, a data protection impact assessment (DPIA) must be carried out.³ Finally, the data collection must be registered by the faculty data manager. The responsible researcher initiates these activities and ensures completion in the correct manner.
- In the case of human subject research, a proposal should furthermore be submitted to the Faculty Ethical Review Committee (FETC) prior to conducting the research.⁴ Data may not be collected until a positive assessment has been received.

Regulations

Storage of data packages

- Research data must be managed in accordance with the applicable university requirements relating to information security, privacy protection and transparency.
- During the research project, data must be stored in an environment with an access restrictions option and back-up facilities. The standard environment provided by the faculty for this purpose is YODA. Other options are the O: drive or SURFdrive. Researchers can consult the Research Data Management Support website to identify a suitable storage environment.⁵ Researchers may propose using a different environment and should first submit a proposal to the faculty data manager.
- If the data contain personal data, the use of YODA as a storage environment is obligatory. It is possible to derogate from this obligation but only with the permission of the Information Security Officer. Local storage is temporarily permitted, for example, because data have been recorded on a peripheral device. However, the data must be encrypted for local storage, and transferred to YODA as soon as possible and destroyed locally. Further guidelines for the storage of data containing personal data follow below.

Handling data containing personal data

- Legal requirements laid down in the GDPR apply to the processing of personal data. Researchers attached to the Faculty of Humanities must therefore process personal data in accordance with the requirements laid down in the GDPR, and in accordance with the requirements laid down by Utrecht University and the Faculty of Humanities. In general, this means that⁶:
 - Before creating a data collection, the researcher must carry out data classification and a DPIA, supported by the faculty data manager or otherwise.⁷
 - Limit: only the personal data required for the research are collected. If, prior to the research, the purpose for which personal data are used to answer a research question is not clear, those personal data should not be collected.
 - Separate: directly identifiable personal data must be separated from the other data and replaced by pseudo codes as soon as possible. Files in which the directly identifiable personal data and the corresponding pseudo codes are recorded, the 'key files', must be separated from the other research data.

³ <https://www.autoriteitpersoonsgegevens.nl/nl/zelf-doen/data-protection-impact-assessment-dpia>

⁴ <https://fetc-gw.wp.hum.uu.nl/en/>

⁵ <https://www.uu.nl/en/research/research-data-management/tools-services/tools-for-storing-and-managing-data/storage-solutions>

⁶ See <https://www.uu.nl/en/research/research-data-management/guides/handling-personal-data>

⁷ <https://www.uu.nl/en/research/research-data-management/guides/handling-personal-data>

- Inform: as far as is possible, the persons concerned will be informed of the research and requested to grant permission for use of their data.
- Provide control: as far as is possible and/or reasonable, research subjects will be given the opportunity to inspect their data, correct their data or withdraw from the research.
- Render unidentifiable: data containing personal data are archived only after anonymisation so that the data cannot be directly traced back to individuals. Anonymisation can be carried out using aggregations, or pseudo codes to replace the personal data.
- Hide: data containing personal data must be stored in locations with encryption.
- The storage or transmission of such data outside the European Economic Area is not permitted.
- Data containing personal data may only be exchanged with third parties provided that explicit agreements (transfer or processing agreements)⁸ have been concluded for that purpose.

Archiving and publishing data

- As soon as (a version of) the data form the basis for a scientific claim expressed in a scientific publication, for instance, the data concerned are packaged together in a data package, archived and published.
- The data package contains, to the extent possible, all the necessary data, descriptions, scripts and images in order to:
 - reproduce the research results;
 - retrieve the origin of the data;
 - understand the collection method;
 - identify the staff involved in creating the data package;
 - understand the objective of the research;
 - know the conditions for reusing the data;
 - know the minimum storage duration.

Depending on the archive, repository or storage environment selected, the researcher can include most of the above information in the standard metadata forms currently available. All the information the researcher is unable to include in the metadata forms should be recorded in a text file that is added to the data package.

- As stated earlier, YODA is the faculty's preferred digital archive and publication platform.⁹ Other data archives, repositories or storage environments may be used. Researchers should consult the Research Data Management Support website to identify a suitable archive environment. Researchers may propose using a different environment and should first submit a proposal to the faculty data manager.
- In principle, data must be anonymised before being archived. Data that cannot be anonymised are archived and published as a data package referenced as 'Closed Access'. Closed Access data packages can only be accessed by the data owners.
- Data to which rights apply, such as copyrights, and therefore cannot be archived for a prolonged period or included in the publication of a data package, must be deleted prior to archiving.
- Data packages are made available as Open Access data to the public, provided that there are no compelling reasons for providing the data under Restricted Access or Closed Access.
- The Faculty of Humanities' preferred user licence is Creative Commons Attribution/Share Alike 4.0.¹⁰

⁸ <https://intranet.uu.nl/verwerkersovereenkomsten-werken-met-persoonsgegevens>

⁹ <https://yoda.uu.nl/>

¹⁰ <https://creativecommons.org/licenses/by-sa/4.0/deed.nl>

- Where possible, the original (primary/rough) data must be archived, together with the scripts used to transform and analyse the data.
- For maintenance purposes and to ensure long-term accessibility, data files should be archived in 'sustainable' file formats, where possible. The Faculty of Humanities uses the DANS list for this purpose.¹¹
- All individuals whose names are recorded in the data package are also identified by a persistent identifier in addition to their first and surname. The Faculty of Humanities uses ORCID as the identifier for individuals¹².
- In principle, data are archived for a period (the retention period) of at least ten years in a data package.
- If data have been archived in YODA, the faculty independently decides at the end of the retention period whether the data should be archived for a longer period and at which location.
- After the data have been archived in a data package, the metadata of the package are published (in other words the metadata are distributed) and the data package is assigned a unique identifier. This occurs automatically in most repositories at the time of archiving. The Faculty of Humanities uses DOI as the persistent identifier for data packages.¹³

III – Data management support

Comprehensive support is available for data management in the form of information on the Research Data Management Support site.¹⁴ Templates for data management plans, standard processing agreements and other documents can be found on the website. Practical support can also be requested for all kinds of topics relating to data management, such as handling data within a consortium and the use of software.

Practical support can also be requested at the Faculty of Humanities. The first person to contact is the faculty data manager.

Support at faculty level

- The data manager: Datanagement.gw@uu.nl
- The Faculty of Humanities Research Support Office (RSO): rso.gw@uu.nl
- Privacy Officer for the Faculty of Humanities and the Faculty of Law, Economics and Governance: Joris de Graaf j.w.a.degraaf@uu.nl
- The Faculty Ethical Review Committee (FETC): fetc-gw@uu.nl

Support at university level

- Research Data Management (RDM) support: info.rdm@uu.nl

¹¹ <https://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/voor-het-deponeren/bestandsformaten>

¹² <https://orcid.org/>

¹³ <https://www.doi.org/>

¹⁴ <https://www.uu.nl/en/research/research-data-management>

- Website: [Research Data Management Support](#)

Links and more information

- [University Policy Framework for Research Data](#) (Utrecht University)
- [Netherlands Code of Conduct for Research Integrity](#) (KNAW, VSNU, NFU, NWO, TO2, VH)
- [Algemene Verordening Gegevensbescherming \(AVG\)](#)
- [General Data Protection Regulation \(GDPR\)](#)
- [Online tool](#) for creating a data management plan (DMP Online)

