

## Faculty of Arts: Research Data Management Protocol and Guidelines

[http://myuniversity.rug.nl/infonet/medewerkers/let/onderzoek/research\\_data\\_management/](http://myuniversity.rug.nl/infonet/medewerkers/let/onderzoek/research_data_management/)

*Please note: this is a copy (Download: 20-11-2017) of the information as provided on the Faculty of Arts Intranet.*

*Some links are available only when you have access to this intranet.*

### Intro: Research Data Management

All Researchers at the Faculty of Arts make use of or create data as part of their research workflow. The data is the evidence that underpins their academic endeavours and thus forms an important part of the academic record. Good Research Data Management (RDM) is therefore one of the prime responsibilities of a professional research organization – not as a goal in itself, but because it is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. <sup>[1]</sup>

The University of Groningen adopted a [university-wide Research Data policy](#) in 2015. The Faculty of Arts endorses the principles of FAIR (findable, accessible, interoperable and reusable) data management. The Faculty Board has established a supplemental faculty research data management protocol. This protocol was developed to help Faculty of Arts researchers meet data management requirements. More information about this protocol can be found in the menu on the left (or: [Faculty RDM protocol](#)).

RDMP web tool

Furthermore, a template for a Research Data Management Plan (RDMP) is developed in the form of an online web tool. This web tool consists of a series of questions that a researcher has to answer. When all answers have been submitted, a Data Management Plan is generated. For more information, see: [RDMP web tool + guideline](#).

Please note: at this moment Internet Explorer (IE) is NOT supported. Please use Google Chrome or Mozilla Firefox.

---

[1] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

## Faculty RDM Protocol

On 11 July 2017 the Faculty Board established the faculty's Research Data Management (RDM) protocol. This protocol asks every academic staff member of the Faculty of Arts – including PhD's – to submit a Research Data Management Plan for each research project in which he/she is the principal investigator (PI). This also applies to doctoral research: primary supervisors are obliged to submit an RDMP for each of their students' PhD research projects within six months of the research project starting. The RDMP will then be added to the PhD student's Training and Supervision Plan (TSP).

### Aim

Research Data Management is a multi-faceted concept that concerns, among other things, the ownership, documentation, standardization, archiving, accessibility and security of research data, and to which ethical, legal and institutional rules and regulations apply. The Faculty protocol is developed to help Faculty of Arts researchers meet these requirements. In addition, a template for a Research Data Management Plan is offered in the form of a [RDMP web tool](#).

Considering that RDM is a relatively new field, with all kinds of ongoing developments, this protocol aims to create awareness of RDM and thus improve the quality of research, including data management. Good data management will help researchers to keep their research safe and secure, make research output more visible and promote collaboration. It will also increase research efficiency: storing and archiving research data properly can save researchers a lot of time in finding and interpreting information as well as the frustration of losing or corrupting it. To this end a [guideline for the RDMP web tool](#) and [support](#) are also provided.

### Context

The Faculty research data management protocol is an elaboration at faculty level on the [UG Research Data policy](#). The Faculty of Arts endorses the principles of FAIR (findable, accessible, interoperable and reusable) data management. The concepts of Open Data and reproducibility, however, are not enforced in this Faculty protocol because they require a much higher level of standardization of and accessibility to the existing digital ecosystem surrounding academic data publications.

Updates to the protocol are expected, as RDM requirements and standards, legal restrictions and IT infrastructure are under constant development.

## **RDMP: what, why & how?**

What is a RDMP?

A research data management plan (RDMP) is a document or a paragraph in your research proposal in which you describe:

- what data you plan to collect during your research project;
- how you are going to store and manage the data during the project;
- how the data will be managed after the project is finished.

Why do you need a RDMP?

A data management plan (or paragraph) makes research data management concrete and practicable: what are you going to do and how?

It saves you time, work and money: by elaborating the management of research data at an early stage, you reduce the chances of having to face unpleasant surprises later in your research.

Nowadays, most funding agencies, such as NWO (The Netherlands Organisation for Scientific Research) or the European Commission (for ERC, H2020), ask for a specific research data management plan as part of a research project proposal.

Also, the University of Groningen has a Research Data Policy that stipulates that it is the researcher's responsibility to draw up and implement a data management plan for all new research.

How do you make a RDMP?

The Faculty of Arts has developed its own RDMP template in the form of a web tool. This RDMP web tool consists of a series of data related questions that a researcher has to answer. When all answers have been submitted, a Data Management Plan is automatically generated:

[Start writing your Research Data Management Plan!](#)

Please note:

- At this moment Internet Explorer (IE) is NOT supported. Please use Google Chrome or Mozilla Firefox.
- Funding agencies may provide a specific RDMP protocol and/or template as part of their application procedure (see for instance [NWO](#) and [H2020](#)). The Faculty of Arts RDMP template is – to a large extent – similar to these specific templates. Filling in the Faculty of Arts web tool thus helps you to come prepared!
- Also, a new version of the RDMP must be created whenever important changes to the project occur due to the inclusion of new data sets, changes in consortium policies or other external factors.

## Storing and archiving

The Faculty RDM protocol asks researchers to archive all data, i.e. raw preliminary data and secondary processed data, underlying a publication, including PhD theses, for at least 10 years in a secure, fixed format in an appropriate University or third-party data service or domain repository – unless the data are already available via public archives or specified otherwise in the RDMP. Both the final publication and a metadata file are archived in addition to the data. The metadata file contains information that ensures that the data is findable, accessible, interoperable and reusable.

Open, unless...

The UG Research Data policy advocates an 'Open unless...' approach. This implies that research data to be archived are made available for access and reuse for academic research in so far as is reasonably possible and with due observance of the proper precautions. The currently available services, structures and expertise, however, are still fragmented, not optimally connected and not easily accessible to all researchers. <sup>[1]</sup>

Therefore, the Faculty of Arts has opted for setting up the Y drive as a basic (minimum) repository for storing and archiving research data, whereby 'archiving' is understood to mean: 'the transfer of material to a facility that appraises, preserves and provides access to that material on a long-term or permanent basis'. The procedure for storing data on the Y drive is described in the [Guideline for the RDMP web tool](#).

It is of course also possible to use other (external) data services or domain repositories. There are hundreds of repositories worldwide, and they may go by different names, such as data centre or archive. Information about data which is retained outside the UG Y drive or institutional repository (in case of GIA) must be registered in the RDMP. A selection of available data services and repositories is provided in the [Guideline for the RDMP web tool](#).

Archiving data does not imply that the data are automatically shared, and sharing data does not necessarily mean open access. The archiving of data allows researchers to meet the obligation to make the research verifiable. There is no requirement, however, to store the data as Open Data: making data accessible also comprises archiving it in a public or non-public place where the data can be made available on request.

Please contact the UG's Research Data Office if you have any questions on internal or external facilities for Research Data Management: [rdo@rug.nl](mailto:rdo@rug.nl)

---

[1] UG discussion paper on a transparent research environment/open science, p. 3.

## Privacy Sensitive Data

Confidentiality needs to be observed when storing and archiving privacy-sensitive data, in accordance with the [Personal Data Protection Act](#), which will be replaced by the General Data Protection Regulation (AVG) on 25 May 2018.

The Netherlands Association of Universities (VSNU) provides an explanation of the Personal Data Protection Act for academic research. <sup>[1]</sup> Their [Netherlands Code of Conduct for Academic Practice](#) includes the important guideline that researchers may only use personal information for research, i.e. for an academic publication. In addition, researchers must use anonymous data as much as possible and minimize the use of directly identifiable information, as far as is justifiable given the nature and purpose of the research and its proper execution.

In addition, participants in research must give explicit consent for their participation, as well as for the generated data being made available to third parties. This makes it important to work with full and correct information letters and consent forms. For information on the procedures to be followed for research involving human subjects, please refer to [CETO: the Faculty Committee for the Ethical Evaluation of Research](#).

The consent form from the original research is leading in the event of the reuse of data that contains personal information. Any permission granted for reuse is solely for the purposes of academic, statistical or historical research, and a work method following the VSNU Code of Conduct for the use of personal information in academic research is mandatory in such cases. This work method must be described in the relevant RDMP.

Personal data must at all times be stored on a secure network and in accordance with legal regulations. Storage, archiving and handling should be described in an RDMP. A privacy impact assessment (PIA) can be performed to map the privacy risks of a project at an early stage and in a structured and clear manner. Please refer to [the UG Research Data Office for more information about a PIA](#).

---

[1] A revision of the code of conduct is currently being worked on, taking the European Privacy Regulations into account.

## **Costs**

Costs relating to data management incurred during or after a research project can be included in a proposal's budget. These may be costs related to temporary storage, to the anonymization or the transcription of data, or to the curation of data before sustainable archiving. Where possible and necessary, researchers should seek to recover the direct costs of managing research data from the research funder.

## **Ownership and breaches**

All research data collected, generated or otherwise acquired by staff and PhD students who hold a permanent or temporary appointment at the UG, or by bursary PhD students who are registered with the UG, are subject to shared ownership by the UG.

In the event of any deviations from this position, for instance as a result of partnerships in research grant projects or contract research projects, a written agreement will record the further arrangements concerning control over the data. In such cases, this agreement forms part of the RDMP of the relevant project.

The stored and archived data must remain in the possession of the relevant research institute if researchers leave the UG; the agreements recorded in the RDMP will remain in force. Departing researchers do, however, retain the right of use: they may take copies of the data with them. It is the responsibility of the direct manager of the departing researcher to ensure compliance with this agreement.

Within the Faculty of Arts, the relevant research institute acts as the legal representative of the UG and has control of all research data collected, generated or otherwise acquired within the institute, unless different arrangements are clearly documented and approved by the institute director. RDMPs and the associated raw and processed data will be provided on request to the chair of the research group, the director of the research institute or the Faculty Board.

Where a breach of protocol results in the unauthorized release of identifiable personal information or non-compliance with external funding or regulatory requirements, this must be reported to the relevant research director (of CLCG, GIA or ICOG) as soon as possible.

## **Evaluation**

The directors of the research institutes promote good practice in all aspects of research, including research data management (RDM). They ensure that staff and students are aware of their responsibilities and obligations in the management of research data, and they ensure adherence to these principles. They promote training where gaps in RDM skills are identified.

The Faculty Board oversees the implementation of and compliance with the UG RDM policy. The present faculty RDM protocol was drawn up to this end. In addition, the Faculty Board provides facilities for RDM, in the form of the RDMP template and the storage space on the Y drive. Staffing has been made available for the further development and management of these facilities.

The current policy will be reviewed within two years (before 1 September 2019) to ensure compliance with current legislation, codes of practice and regulatory standards and to take into account infrastructure developments as well as the needs of the research community. Research integrity, including Research Data Management, is a recurring item on the agendas of the meetings between the institutes' directors and the Dean.

In addition, in 2017 the UG will start an internal audit cycle to evaluate what can be improved upon within the faculty/institute and the UG as a whole.

## **RDMP web tool + guideline**

The Faculty of Arts RDMP template can be filled in via an online web tool:

### **Start writing your Research Data Management Plan!**

Please note: at this moment Internet Explorer (IE) is NOT supported. Please use Google Chrome or Mozilla Firefox.

The RDMP web tool deals with all phases of the Research Data Lifecycle and includes the following sections that are outlined in the guideline (see left menu):

1. [Administration](#)
2. [Data collection](#)
3. [Data ownership and ethics](#)
4. [Data storage and back-up during the research](#)
5. [Data archiving and documentation](#)

Demonstration and Instruction videos

.mp4 Movies. English spoken. Tip: in case your browser does not support the playback of this video, download it and then play it.

[Video: Start to use the webtool for making a Data Management Plan](#)

[Video: How to make a print or a PDF copy of your Data Management Plan](#)

[Video: How to submit your Data Management Plan](#)

[Video: How to review and edit a plan that you already submitted](#)

## **Administration**

The administrative section of the RDMP describes the project and the contact details of the PI responsible for the data, and includes:

1. Name of the researcher and the research institute and research group/degree programme that he/she is part of.
2. Project title, period (start date and expected end date), funding agency, relevant project codes and a brief (3-5 sentences) description of the project and its expected results
3. Person(s) responsible for the project's data management:
  - For PhD projects: PhD student & (primary) supervisor/PI
  - For all other research projects/subprojects: researchers & PI
  - With joint projects with external partners: indicate whether part or all of the data will be collected externally, and who (PI) is responsible for data management at the external institute
4. Version and date of the RDMP, dates of last update of the RDMP.

## **Data collection**

This section of the web tool describes the data that will be collected, generated and used in the project.

### Data guide

There is not one strict definition to describe the wide variety of research data in Humanities research. General definitions include:

- Any data that underpins the production of research output.
- The recorded factual material commonly accepted in the academic community as necessary to validate research findings ([OMB Circular A-110](#)).
- All data generated in the context of academic research that forms the basis of research outcomes and publications

Data may include any or all of the following:

- Text or Word documents, spreadsheets
- Field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Test responses
- Slides, artifacts, specimens, samples
- Collection of digital objects acquired and generated during the research process
- Data files
- Database contents including video, audio, text, images
- Models, algorithms, scripts
- Contents of an application such as input, output, log files for analysis software, simulation software, diagrams
- Methodologies and workflows
- Standard operating procedures and protocols
- Output from statistical packages
- Output from geographic information systems

## **Data ownership & ethics**

This section of the web tool describes whether specific requirements apply with regard to licenses and storage of privacy sensitive data, such as:

- [Medical Research Ethics Committee regulations](#) for research that is medical in nature, or research that requires a medical procedure
- [The Faculty of Arts Ethical Review Committee](#) (Dutch: Commissie Ethische Toetsing Onderzoek, CETO) to provide guidelines and to evaluate research proposals involving human participants.

Also, here you can indicate whether specific arrangements have been made with regard to access to the project data, such as:

- consent forms or other permits or licenses for use or storage of specific data
- copyrights of photos/movies
- contractual obligations to (commercial) external research partners
- embargo requirements for future publications
- specific requirements of funding agencies with regard to data sharing and open access
- specific requirements of scientific journals with regard to data sharing and open access

## **Data storage & back-up**

This section of the RDMP web tool describes how data will be stored and backed-up during the project, and how and where the final data will be archived at the end of the project.

### Storage and back-up guide

For ongoing research, data should be stored securely. Researchers must ensure that the data is backed up regularly. Storage options include the following:

Network storage on the university servers (Y drive): secure and backed up daily; suitable for long-term storage of master data. A folder structure has been created on the Y drive (Y:\staff\let\ResearchData) for the purpose of storing and archiving research data. This structure consists of three layers: 1) research institute or graduate school; 2) research group or centre; 3) personal folder based on staff (p-)number. Each staff member is the owner of his/her personal folder, with a storage capacity of 100 GB. Staff members can determine who has access to their folders and assign rights (via [CIT application](#)). The research director and the coordinator of the research institute have reading rights for all personal folders of the institute concerned.

Cloud-based storage : convenient for easy sharing between collaborators in a project; however, third-party cloud services such as DropBox, Google Drive or Mendeley may not be suitable for storing sensitive data, and their service level agreements should be studied before using them to store your research data. Preferably use the in-house alternative Unishare:

<http://myuniversity.rug.nl/infonet/medewerkers/ict/collaboration/unishare>

Local drives of PCs and laptops : convenient for short-term storage and data processing; not suitable for long-term storage because of potential hardware failure, theft & loss.

External storage devices , e.g., external hard drives (NAS), USB memory sticks, CDs & DVDs: cheap and portable, but not suitable for long-term storage or back-up: the longevity is uncertain, and these devices are easily damaged, lost or stolen.

## Data archiving & documentation

This section of the RDMP web tool describes how and where the data is archived at the end of the project, how the data archive is organized in terms of metadata, which data formats are used and how these are accessible. Please indicate which additional software is required, with links to source sites if relevant.

### Archiving guide

Upon publication of research, the publication is registered in PURE, preferably via [Green Open Access](#), and all data underlying the publication are archived, whereby archiving means here: to store the data, including metadata, in a fixed form on a secure system that ensures data accessibility for the required period of 10 years. All primary and secondary data should be archived in a single zipfile, including the final publication and at least one metadata file that documents the data archive.

The following options are available for archiving your data:

- [GIA respository](#) (for GIA data only).
- **Y drive (University network)**. A folder structure has been created on the Y drive (Y:\staff\let\ResearchData) for the purpose of storing and archiving research data. This structure consists of three layers: 1) research institute or graduate school; 2) research group or centre; 3) personal folder based on staff (p-)number. Each staff member is the owner of his/her personal folder, with a storage capacity of 100 GB. Staff members can determine who has access to their folders and assign rights (via [CIT application](#)). The research director and the coordinator of the research institute have reading rights for all personal folders of the institute concerned.
- [DataverseNL](#). DataverseNL provides archives for the faculty's three institutes and enables researchers to provide access to the data (up to 5 GB) they have stored in a user-friendly way. Dataverse uses persistent URLs, making data permanently accessible even after it is moved to a different location. Please note: DataverseNL is not suitable for privacy-sensitive data! To start using DataverseNL, please send an e-mail to: [researchdata@rug.nl](mailto:researchdata@rug.nl)
- **Other means of archiving**: It is of course also possible to use other (external) data services or domain repositories. There are hundreds of repositories worldwide, and they may go by different names, such as data centre or archive. Information about data which is retained outside the UG Y drive or institutional repository (in case of GIA) must be registered in the RDMP.

### Metadata guide

**Metadata definition**: Documentation or information about a data set. It may be embedded in the data itself or exist separately from the data. Metadata may describe the ownership, purpose, methods, organization and conditions for use of data, technical information about the data, as well as other information.

Metadata provides information about the data collected in a project, allowing the data to be found, identified and interpreted after a project is finished. Each data archive requires a description of its contents; large archives consisting of multiple folders also require a descriptive metadata file for each folder. A metadata file is a simple text file named read\_me\_first.txt including at least the following information:

- The description/reference of the publication (for ReMa students: include your staff number; for publications: include the doi).
- An overview of the contents of the zip archive, and how this relates to the publication: indicate how the data was processed, which data was used in which chapter and to produce which results/plots.

- Which computer programs were used and what version of that program and links to source sites where relevant.
- If relevant: a short description plus links for data that are archived externally.
- Contact details of the author(s)/data owner(s).

The preferred way of data storage is as tab or comma-delimited text files with variable names in the first line, with an associated R script that reads the data file, as this makes data robust towards future changes in software and data file formats. For similar reasons of compatibility and future accessibility, consider using the file formats in the May 2013 [KNAW-DANS Preferred Formats](#) overview, for other data types.

All file names, metadata and other descriptive files, as well as comment lines in code must be in English to ensure general accessibility of the data. File names should not contain special characters other than dashes (-) or underscores (\_); use underscores instead of spaces. Preferably, use the naming standards defined in the file naming guide below.

#### File naming guide

The naming of data is a crucial part of data management. Naming the files with a sound logic allows for easy identification of ownership. In addition, it enables the fast connection of a given datafile to the respective documentation. Examples:

- minimalistic:
  - Name\_yyyymmdd\_Keyword
  - Name\_Initials\_yyyymmdd\_Keyword
- extended:
  - FamilyName\_GivenName\_yyyymmdd\_Keyword

Please use separate sub-folders for different research projects!

## **Practical support**

Please contact the [Research Data Office](#) for advice and support regarding:

- Drafting a Research Data Management Plan for funding proposal
- Legal aspects
- Data description
- Technical facilities
- Drawing up cost budgets

Please contact Christina Englert for technical aspects of the faculty RDMP web tool: [c.englert@rug.nl](mailto:c.englert@rug.nl)

Please contact Dries Gankema (demand manager) if you would like to request extra space on the Y drive . Up to 3 terabytes of extra space can be requested free of charge: [a.j.gankema@rug.nl](mailto:a.j.gankema@rug.nl)

If you would like to share your personal folder on the Y drive with a UG colleague please use the [CIT web application](#).