# Research Data Management Policy
within the Faculty of Arts (FoA)

Update October 2022

## 1. Introduction and context
Investing in the practice of transparent and reproducible science is a high priority of all academic institutions worldwide, organisations like NWO, KNAW and international research funders in general. Research data management, personal data protection and open science are pillars of this movement.
This research data management (RDM) policy provides context and guidance to all members of the Faculty of Arts (FoA) that deal with research data in its broadest definition. The document will be regularly subjected to a review as it is affected by changes in the various regulations that underpin it.

The update of this document is based on the [UG Research Data Policy](#) dated June 2021, and needs to be read alongside the following documents:
- [the European General Data Protection Regulation (2016)](#)
- [Ethics and data protection (2018)](#)
- [the 'Practical Guide to the International Alignment of Research Data Management' by Science Europe (2021)](#)
- [the Netherlands Code of Conduct for Research Integrity (2018)](#)
- [the 2020 University of Groningen Code of Conduct on Integrity](#)
- [the University of Groningen PhD regulations (2022)](#)
- [the Faculty's policy and procedures on research ethics (CETO)](#)
- [the ethical code of the National Ethics Council for Social and Behavioural Sciences (NETHICS)](#)

## 2. Organisation and responsibilities
There is joint responsibility for research data management within the UG. The guiding principle here is the Higher Education and Research Act and the elaboration thereof in the Administration and Management Regulations of the UG, which allocates the responsibility at different levels.
Responsible data management requires collaboration between management (organisational units within the university), scientific staff and support staff.

**Faculty Board**
The faculty board has a duty of care for its researchers and (PhD) students. They are responsible for setting up, maintaining and overseeing the implementation of a responsible data management policy in accordance with the UG Research Data Policy.
The Board of the FoA has approved this update of the RDM policy document in October 2022. Changes to the RDM policy will also be subject to the approval by the Faculty Board.

**CETO**
The Faculty of Arts (FoA) committee for ethical review of research projects called [CETO](#) (*Commissie Ethische Toetsing Onderzoek* in Dutch) evaluates research projects within the Faculty (and sometimes from UG researchers outside the Faculty) involving human subjects and other ethical aspects. CETO is accountable to the Dean, who acts as Placeholder of Research.
Since July 2020, obtaining ethical approval is obligatory for researchers and PhD students within the FoA who run projects that involve (data of) human subjects.

**Graduate Schools**
The Graduate Schools are responsible for the proper onboarding, training and guidance of PhD students. In this role they can also point out the responsibility of PhD students regarding data management and the attention they should pay to research data management planning (RDMP) in their training and supervision plans (TSP). For training on research data management and data protection principles the Graduate Schools collaborate with the Digital Competence Centre (DCC). The Graduate Schools also play a role in preparing PhD supervisors to make arrangements, together with the PhD student, for the PhD student's research data to be made reusable.

**Support**
The FoA's Research Policy Officer coordinates the formulation and implementation of the research data management policy. The Policy Officer works closely together with the Data Steward/CETO secretary, and the FoA's Privacy & Security Coordinator on matters concerning the GDPR.
The data steward/secretary CETO monitors all incoming registrations of research projects and contacts researchers when research is deemed eligible for ethical assessment and/or in case of possible (privacy) risks. Together with the Privacy & Security coordinator, the data steward/secretary CETO performs a Data Protection Impact Assessment (DPIA) if deemed necessary. Additionally, the DCC offers university-wide support on research data management and protection and related IT solutions for both researchers and support staff.

**Researchers**
Three research institutes accommodate the research of the Faculty of Arts, namely the Centre for Language and Cognition (CLCG), the Groningen Research Institute for the Study of Culture (ICOG) and the Groningen Institute of Archaeology (GIA). Each institute consists of smaller centres or research groups. It follows that the FoA houses a mixed company of researchers, working with many different categories of research data or in fields that are not strictly data-driven.
And yet, this RDM policy applies to all researchers and PhD students of the FoA, ensuring that the faculty is exercising its oversight responsibility.

# 3. Principles of Data Management

**Scientific Integrity**
There are several key overarching principles that govern research at the University of Groningen and scientific research in general:
1. All research is to be undertaken according to relevant legislation, the requirements of ethically responsible research, such as informed consent, academic integrity, the arrangements set out in cooperation agreements, and the research conditions of funders.
2. When personal data are processed in research (within or from the EU), these must be handled in a way that meets the European General Data Protection Regulation (GDPR). In particular, this entails the requirement to demonstrate that appropriate technical and organisational measures to protect the identity of research participants are in place.
3. Where feasible and ethically appropriate, researchers will make their data FAIR (see below).
4. Research methodologies, experimental design, data, data analysis, statistics, results and interpretations should be subject to peer review and scrutiny, and all researchers should be able to substantiate their choices, and be able to provide access to all research results (papers, data and code) for validation.

**FAIR research data**

The UG RDM policy is based on the FAIR data principles included in The Netherlands Code of Conduct for Research Integrity. These four principles are:

Findable
- Data is sustainably stored and curated
- Data is provided with metadata, including affiliation and, if possible, with a persistent identifier
- Data is registered on the basis of metadata in the [research database](#) of the university

Accessible
- Data is accurate, complete, reliable, authentic and provided with metadata and, if possible, with a persistent identifier
- Data is available for checking and further research after completion of the research and/or the departure of the researcher
- Data is openly available, unless valid reasons prevent this

Interoperable
- Data is stored in sustainable file formats
- The discipline specific standards for the enrichment of data are applied
- Data (and metadata) is provided with references to other relevant material

Reusable
- The origin of the data and the affiliation of the researcher involved is clear (Provenance)
- The discipline specific standards for data management are applied
- The conditions for use are clearly described

There may be sound reasons not to make data FAIR, however. This policy is based on a "comply or explain" principle, meaning that exceptions to this policy are possible. A researcher may wish not to make the data publicly available when:

1. Making the data publicly available would contradict any agreement or undertaken given as part of the informed consent process
2. Where public release might have negative consequences for the people researched
3. When there is an obligation or objective to protect results so that they can commercially or industrially exploited (or reasonably expected to be exploited)
4. Where public release could represent a security concern
5. If open access would jeopardise achievement of the main aim of the research
6. If there is any other legitimate reason not to publish the data.

**Open Science and Open Access**

Open Science is one of the focal points of the Strategy Evaluation Protocol (SEP). The four pillars of open science are open data, open code (research software), open papers (open access) and open reviews (open evaluation). The FoA stimulates making research data, lab notes and other research processes freely available where possible, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods. Doing so will not only result in a more transparent, verifiable and faster research process, but will also enable companies and societal organisations to readily gain access to, and use scientific information in an early phase.

# 4. Research data management workflow agreements

This Research Data Management Policy formulates a set of agreements that apply to all research conducted within the FoA and that refer to different stages of the research process. The FoA Research Data Management Protocol provides a practical guide listing best practices and preferred solutions.

## Research data before research

*Registration*

Each researcher (or group of researchers) has to draw up a Research Data Management Plan (RDMP) for each research project during the phase of research design (for example as part of an application for research funding). This requirement applies to all research projects in which research data are being collected or reused. All PhD students are obliged to draw up and hand in an RDMP at the start of their research project, as part of their Training and Supervision Plan (TSP).
The FoA offers facilities for digitally submitting an RDMP according to the FoA template (see the FoA RDM protocol).

*GDPR & ethical assessment*

Researchers and PhD students have to act in accordance with the General Data Protection Regulation when dealing with personal data, and in accordance with other statutory and ethical rules (where they apply). In addition, obtaining ethical approval from CETO is obligatory for researchers and PhD students within the FoA who plan for research projects that involve human subjects or human subject data.

## Research data during research (storage)

In the life cycle of research data it is important to distinguish between the phase in which data is still being manipulated and analysed (mutable data) and the phase in which research data has been processed and become static (immutable data). The same distinction exists for storage (during research) and archiving systems (after research), which have different functional and technical requirements.

*On-premises storage*

On-premises storage refers to storage that is hosted by the UG itself in its own data centres, managed by the Centre for Information Technology (CIT), and that is regularly backed-up. Examples are the X- or Y-drive attached to the university workplace (UWP). Storing your research data on-premises is always preferred over using personal (external) storage devices or cloud storage, for reasons of data security and protection (prevention of data loss and data leak).

*Cloud storage*

Many third-party (non-UG-hosted) systems or applications are available to researchers to help collect, share, analyse and/or visualise research data (e.g. Qualtrics, ArcGIS Online, Google Drive). If third party tools process, hold or retain research data in any form in the course of a research project researchers should specify their use of these systems in their RDMP. The FoA strongly recommends researchers to move their research data to on-premises storage as soon as possible after the collection and/or processing of data is complete and after collaborations on shared (cloud) storage have been concluded.

## Research data after research (archiving)

*Archiving and retention*

Data gathered for scientific, historical or statistical purposes should be stored for a period of at least ten years after it has been processed (to enable scrutiny in relation to concerns about matters of scientific integrity).

Data can be kept longer than ten years for historical, statistical or research purposes (for example in the case of archeological or historical data).

Data must be archived such that:
- an academic staff member from the same discipline can understand and verify that data, and possibly replicate the research (data should be accompanied by relevant documentation and metadata);
- it is accessible in the long term, i.e. stored in a future-proof format.

Non-digital data must be duplicated in digital or non-digital form, so that both the researcher and the Faculty can keep a copy. If it is not possible to make a duplicate, then the data must be stored at the Faculty whereby the researcher will have unrestricted access, including after they leave the employment of the University.

*Personal data storage limitation*
In case of personal data, the GDPR specifies a storage limitation principle: "Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed." The GDPR does not set specific time limits for different types of data, but requires that controllers and processors set limits based on the purposes of the processing or that de-identification measures are taken. Researchers should therefore specify when they plan to remove the personal data they have collected, or what (de-identification) measures they have taken to be able to retain their data, in their RDMP.

**Research data publication**
Research data publication ranges from making data findable (registering the availability of a dataset externally) to making data completely and openly available for reuse. Any dataset can be registered in the UG Research Portal (using PURE) so that it is findable externally. Additionally, researchers can upload archived research data to an external repository.
A data repository is a place that holds data, makes data available to use, and organises data in a logical manner. A data repository may also be defined as an appropriate, subject-specific location where researchers can submit their data. The FoA strongly encourages researchers to deposit their data and software in an appropriate data repository to enhance their findability, accessibility and impact.
Following the FAIR adagium "as open as possible, as closed as necessary" the researcher applies appropriate technical and organisational measures to protect the data, such as anonymization or pseudonymization (in case of personal data).

**Research software publication**
Research software is defined as any piece of code or script that enables researchers to process, manipulate, generate and analyse data or automatize or test these procedures. The availability of software developed for the purpose of producing research results, together with the respective research data, is essential for the reproducibility and transparency of research. The FoA strongly advises researchers who work with research software to seek the help of the DCC for drafting a Research Software Management Plan in the research design phase, in preparation of software publication after research.

**End of UG affiliation or employment**
Researchers who leave the FoA are obliged to deposit and describe the research data they have collected during their work at the University of Groningen. The line manager (chair holder) monitors the correct and complete deposit of the research data.
PhD students are obliged to properly deposit and describe their research data at the end of their research project (or when leaving the FoA prematurely). The GSH coordinator (in consultation with the supervisor) checks whether all relevant data has been deposited. The PhD defence ceremony will not be scheduled if this requirement has not been met.