Table of Contents

Research Data Management Policy of the Faculty of Science and Engineering

October 2022

1. Introduction

This document describes the general Research Data Management Policy for the Faculty of Science and Engineering that applies to all its Research Institutes and Research Schools to which FSE is affiliated. It outlines the procedures expected from all persons (including academic staff, support staff, PhD candidates, Bachelor and Master students, guest researchers) who generate data as a result of performing research at the Faculty.

The Research Institutes within the Faculty of Science and Engineering act in accordance with the Faculty RDM policy. Further information regarding implementation of Faculty policy and Institute-specific policies related to research data management can be obtained from the individual Research Institutes.

2. Governance and standing of this document

This policy document was developed by the members of the RDM Workgroup with input from the Research Institutes, and has been subsequently approved by the Faculty Board. This document may undergo revisions in the future subject to changes in the various regulations that underpin it. The following documents therefore also apply:

- the University of Groningen Research Data Policy (Revised 2021)
- the Netherlands Code of Conduct for Research Integrity (2018)
- the Netherlands Code of Conduct for Scientific Practice (2012)
- the European Code of Conduct for Research Integrity (2017)
- the European General Data Protection Regulation
- the Faculty's policy and procedures on Research Ethics (2023) (under development)

3. Principles of Research Data Management

Research data management is a term that describes the organization, storage, preservation, and sharing of data and samples collected and used in a research project. For the purpose of this document, the Faculty of Science and Engineering defines data as any form of **digital data** generated in research (raw primary data, metadata, processed secondary data, recordings, digital lab journals, etc.) including **code and software** developed in a research project, and **non-digital data** including **physical, chemical, and/or biological samples or products** of research.

FSE aims to contribute to science and society and therefore endorses the principles of Open Science following the FAIR principles (Findable, Accessible, Interoperable, Reusable). Data generated and collected in research, including research conducted as part of educational programmes, should therefore meet the following criteria as defined by the Board of the University of Groningen:

- are accurate, complete, reliable, authentic, documented and provided with metadata;
- are safely stored with minimum risk of loss and for at least 10 years;
- are made openly available for review and further study after completion of the research and/or departure of the researcher, unless ethical, legal or contractual obligations prevent this.

Published research results must clearly show:

- upon what the data and the conclusions are based;
- how the data are collected and the conclusions are derived from them;
- where and how the data can be found and verified.

4. Guidelines within FSE

All data generated and collected in research projects within the Faculty of Science and Engineering[1] need to be properly stored and archived for purposes of preserving valuable datasets, software, and samples or products of research, and for reproducing and verifying research results (as a matter of scientific integrity).

The type of research data, software, and samples generated and collected within the Faculty will vary due to the multidisciplinary nature of the Faculty, see section 3 for definition of data. Examples of research data include, but are not limited to, observational data, experimental data, algorithms, software code, simulation data from models, or processed data. Data types

[1] All research data generated within the Faculty of Science and Engineering remains property of the University of Groningen.

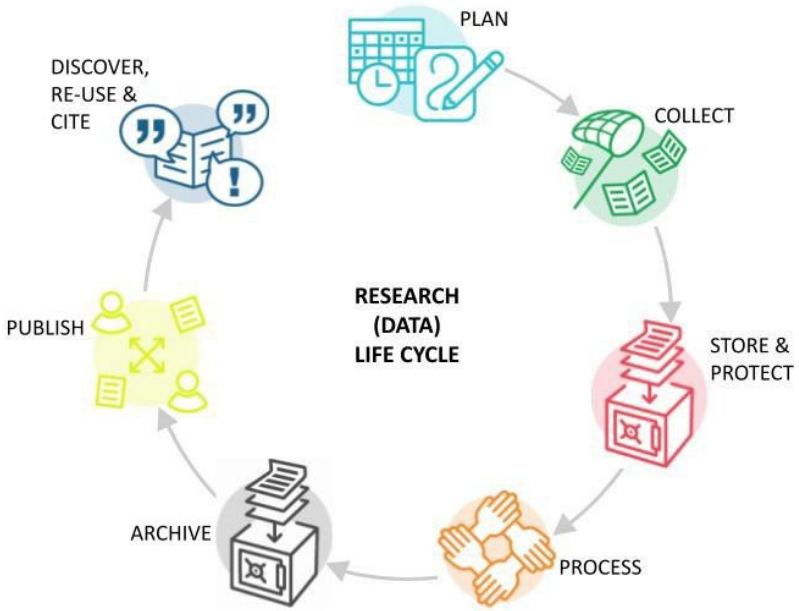
could include biological or chemical samples, text, numbers, images, 3D models, lab or field journals, software, audio files, video files, reports, and so on.

All persons generating research data will need to decide per research project what type of data will be collected and stored and what (file) format will be the most appropriate for storage and archiving. This information will be included in the projects **Research Data Management Plan** (see section 4.1).

4.1 Research Data Management Plan

A Research data management plan (RDMP) is an official document that describes how primary (raw) and secondary (processed) data are collected and used in a project throughout the **research lifecycle**, including where and how the data are stored and archived, who owns the data and other legal issues, and who has the final responsibility for the project's data collection, processing and archiving at the end of the study.

The research life cycle (image retrieved from the DCC website)



All research projects carried out within, or affiliated with, the Faculty of Science and Engineering will be covered by an RDMP. Requirements for an RDMP are:

- An RDMP will be written and filed before or upon the actual start of the project to ensure that all research data that will be generated for the project will be stored in the correct format and categorized properly.
- A new version of the RDMP will be created whenever important changes to the project occur due to inclusion of new data sets, changes to existing data sets, changes in data handling, changes in consortium policies or other (external) factors.
- All RDMPs are to be stored in a webtool and may be printed/saved as pdf for additional purposes.

Research Institutes will provide students and researchers with a template (in a webtool) for filling out an RDMP according to the best practices of the Institute which meets the guidelines of FAIR principles (further outlined in section 4.2). Any accompanying documents,

permits and agreements related to the RDMP (including special considerations, see section 4.3) should also be documented in the RDMP.

4.2 Data collection, storage and archiving according to FAIR principles

All **digital forms of (meta)data**, including software and digital lab journals, are collected, stored, and archived in UG storage facilities that guarantee access and retention of data during a project and after a researcher leaves the University (see **Appendix 1** for a list of suggested storage facilities).

Note: Any raw, primary (meta)data that is collected and stored at an external institute falls under the responsibility of the external institute and does not need to be deposited on UG network storage. All processed, secondary data such as spreadsheets, databases, models, scripts, code etc. must be saved in accordance with the RDMP and stored in UG storage facilities.

Any data stored in (inter)national databases must also guarantee the same access and retention of data during a project and after a researcher leaves the University. Solutions may be needed for exceptionally large data files, in which case the RDM policy of the Research Institute should provide instructions.

All **non-digital forms of data,** such as biological, chemical or physical samples, or hand-written lab or field journals, are stored in UG facilities that guarantee retention of and access to the data. The data must be explained in the RDMP indicating the contents of the data, storage information (if applicable) and location where the data can be found. (see section 4.3 *Special considerations*).

4.2.1 Data formats (*Findability* and *Interoperability*)

Whenever possible, both data (including code) and metadata should be stored in an open data/open source code format, clearly named and easily **findable** (using digital identifiers, appropriate file names, etc.). The institute RDM policy provides guidelines for the naming conventions used within the institute.

4.2.2 *Accessibility* and *Reuse* of data

RDMPs and associated raw data are made available upon request. The RDMP for the research project should detail under which restrictions the data will be available and what limitations (ethical, legal, or otherwise) prevent access to data if applicable.

4.2.3 Retention period

All forms of data must be saved and archived for a period of **10 (ten) years**, commencing from the formal end date of the project, but at least 10 years after the last publication resulting from the project.

In cases where sensitive personal data are generated (see section 4.3.3 below), data is retained according to GDPR regulations (*the period for which personal data is stored is no longer than necessary for the task performed*.)

4.3 Special Considerations for data

4.3.1 Software as a data type

If software will be used or developed to analyze, simulate or model data and automatize data processes during a research project, it should be kept and documented following the best practices of reproducible workflow. If the software will be made Open, then the repository should contain the appropriate license and the citation file. A research software management plan (RSMP) should be developed as part of an RDMP. The RSMP includes:

- measures that will be taken during the project to ensure long-term sustainability of the software developed in the project (maintenance).
- measures that will be taken to support the software after the completion of the project.
- the resources that are needed to ensure the long-term usability and availability of the software, and how these resources are funded or obtained.
- information on where and how the software code is stored and how access can be obtained
- See **Appendix 2** for information on creating an RSMP.

A reproducible and reusable research software should comply with following best practices:

- Use of code versioning and sharing via a code versioning platform
- A README file that clearly describes the purpose of the software, its version, how to install and run it, its requirements and the contact person.
- License
- Citation file

For further information on reproducible workflow please look at the research software management section of the DCC website.

4.3.2 Biological, chemical or physical samples or products of research:

If biological, chemical or physical products are generated as part of a research project, the following information should be registered:

- name of collector, date, and time of collection,
- documentation of storage contents, location, and storage and handling conditions (such as temperature, light/dark, thawing times, etc.),
- special safety measures or considerations (GMOs, chemicals, radioactive materials)
- any permits and/or (national/international) regulations that apply.

4.3.3 Working with personal data

Personal data refers to any information that relates to an (in)directly identifiable living individual. Research data generated may contain personal data, for instance, when research involves collecting data about human subjects (such as research involving human participants or collecting data from users). Processing of personal data that poses a high risk to data subjects requires a Data Protection Impact Assessment (DPIA). A high risk can be caused by:

- the nature of the data (for instance specialized data or sensitive personal data),
- the intended purpose of the processing (will the data be used for profiling, or making automatic decisions),

- the procedures used (such as combining files or large-scale monitoring, or developing new technology),
- the possible consequences (for instance no longer being able to exercise a right, or losing a contract or service).

If a research project involves collecting any personal data, *you are required to contact the FSE privacy and security committee* <privacy.FSE@rug.nl> to discuss and file a DPIA for the project.

Before completing an RDMP for the project, you are asked to contact the UG Digital Competence Center (DCC) <dcc@rug.nl> for advice. The advice of the DCC should be incorporated into the first draft of the RDMP. The first draft of the RDMP should include the DPIA and be completed before the start of the project.

Note: Personal data can only be stored on UG storage facilities and processed on UG computing facilities, or, in the case of collaborative research projects, in suitable storage facilities of the leading research institution.

4.3.4 Ethical Permission for research – (FSE Ethics document under development)

Any research involving the use of human subjects (medical, non-medical), development of medical devices, use of animals, or use of algorithms with potential impacts on individuals or the human population (such as harming the state or integrity of a person as a consequence of data collection or data processing by means of algorithms) should first be approved by an ethics committee to ensure that all ethical considerations and safety measures are properly followed:

- All research involving the use of animals is covered by the *Instantie voor Dierenwelzijn* (IVD), contact <ivd-fse@rug.nl>.
- All medical-related research is covered by the *Wet Medisch-Wetenschappelijk Onderzoek met mensen* (WMO). Contact the Medical Ethics Committee (METc) at the UMCG. Information can be found on the METc website.
- All non-medical related ethical approval requests are handled by the (to be installed) FSE Ethics Committee.

All requests for ethical approval should be initiated before the project starts. This is to avoid any delays in the project (such as collecting data, submitting applications to funding agencies or publishing scientific results). Where research involves collecting personal data, a request for ethical approval should be completed immediately after a DPIA has been filed.

Accompanying documentation:

- FSE Ethics Committee (under development), contact a.brandt@rug.nl for ethics related requests.
- See the DCC website for more information on ethics in research.

4.3.5 Other special considerations

All other special or legal considerations or obligations in collecting research data, storage of data and archiving should be covered in an RDMP (see DCC website).

4.4 Data Transfer Agreements

A Data/Material Transfer Agreement (DTA or MTA) is needed whenever data or materials are transferred between parties outside the UG. A D/MTA identifies the involved parties, details the allowed use of the data, details the ownership of the data, and details special requirements for transmission, storage, use, and publication of the data. As per the normal procedures and/or arrangements in the research institutes, all concept DTA/MTAs should be approved by ABJZ (UG legal department) before being signed by the FSE Managing Director. See the legal department website for guidelines on what should be considered when collaborating externally.

4.5 RDMP Compliance

The collection and storage of data as described in the RDMP should be monitored throughout the research lifecycle (refer to figure in section 4.1) to ensure compliance (for the purpose of verifiability) and adherence to the FAIR principles, unless ethical, legal or contractual obligations prevent this.

Data collection, storage and processing of data during a research project

All persons who produce or are involved in data production must comply with the RDMP of the Research Institute in which the research is executed. Compliance is safeguarded by the head of the research unit. Research data management should be regularly discussed in the Institute. For staff and PhD students, at least during the annual R&D interview.

Archiving data after project completion or publication:

- For intended publications in any form, the student or researcher should compile a documented archive of all data underlying a publication (including publication 'early online') and store it in accordance with the RDM policy of the Research Institute.
- All data collected in the context of a BSc or MSc research project should be deposited no later than the date of handing in the final version of the BSc or MSc-report.
- All data collected in the context of a PhD study should be deposited in agreement with the first promotor/daily supervisor upon deposition of the final thesis manuscript in HoraFinita, including parts/chapters that have already undergone a previous data storage procedure i.e, as (part of) a publication.
- If a researcher leaves the University or the research group before the end of the project, all data should be deposited and handed over to the supervisor or head of the research unit.

At all times:

RDMPs and associated research data are to be made available to the Board of the Research Institute, the Board of the Faculty of Science and Engineering, the Executive Board of the University, or other relevant authorities upon request.

4.6 Destruction of research data

The Research Institutes are responsible for further retention or destruction of research data after the mandatory 10 (ten) year retention period in line with the policies of the Institute. (An exception is the retention and destruction of sensitive personal data, see section 4.2.3).

4.7 Communication and Training

All persons working with data (BSc, MSc, PhD candidates, scientific and support staff members) will receive information and training on research data and software management (for example, as part of their educational programme or onboarding process) and should be continually updated on new developments in RDM.

The DCC provides training on research data management.

5. Responsibilities

Faculty Board

- The Faculty Board is accountable to the Executive Board of the University of Groningen for research data management policy of the Faculty.
- The Faculty Board will ensure that the Research Institutes comply with the UG and Faculty RDM policy.
- The Faculty Board will assist the Research Institutes in maintaining good RDM practices (such as employing data stewards).
- The Faculty Board is responsible for communicating any changes in the university's or faculty's research data management policy to the Research Institutes.

Research Institutes

- will ensure that the institute has its own Research Data Management policy that extends both the university's and faculty's policies.
- will ensure that a Research Data Management Plan template is available within the institute.
- will ensure that the institute's own Research Data Management policy is publicly available on the UG's website.
- will ensure that the RDM policy of the Faculty and Research Institute are communicated to staff and that staff are made aware of the services and support available.
- will ensure that the RDM policy is being followed within the institute and that all data has been properly stored and archived after the end of a research project.

School Science and Engineering and GSSE

- will ensure that BSc/MSc students and PhD candidates are informed about training on research data management and provide appropriate credits for undertaking such training.
- will ensure that students have made proper agreements with their project supervisors on storage and archiving of data at the end of the research project.

- For PhD candidates, the GSSE ensures that writing an RDMP is part of the obligatory training and supervision plan (TSP).

Principle Investigators/Group leaders

- will ensure that an RDMP is generated for all research projects (including BSc/MSc, PhD, postdoc, and visiting researcher projects) according to the RDMP template of the research institute,
- will ensure that RDMPs are continually updated and adhered to by all project members,
- will ensure that students and staff who have performed research projects in their group have properly stored and archived their data,
- will ensure that the RDMP aligns with requirements of funding agencies, research consortia, private parties or any other third parties,
- will ensure that a data and material transfer agreement (D/MTA) is in place for any data transferred outside of the University.

All persons involved in research

- will ensure that research data, code and any other materials (biological, chemical, or physical products of research) needed to reproduce research findings are properly stored according to FAIR (Findable, Accessible, Interoperable and Reusable) principles which guarantees that data will be available for at least 10 years from the end of the research project, and made available upon request.
- will ensure that an RDMP is in place which covers details on the types of raw and processed data that is generated (including storage) in their research project(s).
- will ensure that the relevant metadata is stored in a suitable repository.

Appendix 1 - Examples of Acceptable forms of digital storage media:

Refer to the DCC website for more information on acceptable forms of digital storage media.

Appendix 2: Example of Research Software management plan (RSMP)

Depending on its impact on the research results, research software can be classified into three categories: low, medium and high impact (mission critical) software. For the definitions and examples of these categories please look at the research software management section of the DCC website. Based on the category different types of measures need to be taken to ensure the reproducibility of the results and the reusability of that software.

A research software management plan should involve in minimum the following three points:

- Purpose
- Reliability
- Maintenance

See National Guidelines for SMPs.

Contact the DCC for further help with developing your RSMP