# Research Data Management Policy

within FEB/SOM (work in progress)

Update March 2022

Status:
For advice IRB: January 27
For advice SOM' scientific committee: February 15
For approval Faculty Board: Early March
To the University Board: March

## 1. Introduction and context

The Board of the University of Groningen has requested all institutes to update their Research Data Management Plans (RDMP) dated 2017 for their own institute/school. Data management in the broadest sense is high on the agenda of not only the University of Groningen, but also of other universities in the Netherlands and abroad, and organizations like NWO, KNAW and the EU. To anticipate current and future developments, it is advisable not to have rules carved in stone. Therefore, this update still is a policy document in progress that will be discussed and updated on a regular basis.

The update of this document is based on the UG Policy Document dated June 2021. UG policies are 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

The UG has formulated four points of attention for the upcoming years:

- The monitoring of data management plans
- The training of researchers and those providing support to the research
- Research data in education
- Sharing good practices

## 2. Governance and Organization

### Faculty Board

This update of the RDMP policy document has been approved by the Board of the Faculty of Economics & Business in [month, year]. Changes to the RDMP will also be subject to the approval by the Faculty Board.

### Institutional Review Board

There appeared to be much overlap between issues related to ethics, data management and privacy. In the summer of 2018, we therefore merged SOM's ethics committee with the data management committee into a new committee named the Institutional Review Board (IRB). With respect to data management, the responsibilities of the IRB are to discuss developments in the area of data management and keep policies and policy documents up to date. With respect to research ethics, the first task of the IRB is to provide advice to researchers, research programmes and the Vice dean of Research about matters concerning ethical conduct in research. These may be issues related to, for example, research integrity and the use of human participants in research. The second task is to evaluate ethical aspects in research proposals by staff that involve new data collection among human participants for which a SOM staff member is the principal investigator. These proposals have to be submitted to the IRB before data collection takes place, and this rule applies to all data collection efforts by students, PhD candidates and faculty alike. The IRB is accountable to the Vice dean of Research.
Per February 1, 2022 the composition of the IRB is as follows

Prof. Adriaan Soetevent, chair
Dr. Sumaya Albalooshi
Prof. Jutta Bolt
Dr. Joost van de Brake
Dr. Florian Noseleit
Ms. Rina Koning (policy officer research), secretary
Mr. Jan Willem Oordt Msc (legal department), advisory member

### Support

Data management policies and activities are coordinated by FEB's policy officer research ms Rina Koning. She is also FEB's Privacy & Security Coordinator as well as secretary to the IRB. Because of the overlap between the different areas it is efficient to combine these tasks in one person. FEB's funding officer Ms Astrid van der Veen-Mooij also provides support especially related to grant applications. She also acts as backup. The SOM office provides administrative support.
Within the university, but also at the national level, much is going on in terms of organisation and infrastructure. Relatively new are data stewards, but often with very different job descriptions per organisation. For example in NWO funding applications the participation of data stewards is

required. Another development is the introduction of the Digital Competence Center (DCC). It is not clear yet to what extent and what type of support will and can be offered and therefore it is difficult to say what FEB's future needs will be in terms of support.

## FEB data policy

This RDMP, including rules/regulations, is applicable to all staff and PhD students of FEB. Each researcher or group of researchers has to fill in a Personal Data Management Plan provided by SOM. A tool has been developed to make it as easy as possible for individuals to provide all the required information including ethical issues about their research projects. Because, as of 25 May 2018, all organizations and institutions in the European Union have to be compliant with the General Data Protection Regulations (GDPR), the RDMP tool also includes all relevant and required privacy information.
An extension of the tool will be considered. We will explore the option of adding information in a later stage like the official end date of the project, changes to the original project and information about the storage of data (where) and publications (DOI). Another question would be – line with the the Open Science Programme document – whether plans should be published.

## PhD regulations

In the PhD regulations it is stated that
*In order to safeguard the storage of the thesis and to ensure that the thesis and the research data remain available to the University of Groningen for further research and education, the PhD student must sign the license agreement, as included in Appendix 8. This agreement must be signed before the thesis is submitted for approval by the Assessment Committee. Insofar as the University of Groningen is the employer of the PhD student, the CAONU and the Dutch Patent Act apply with regard to intellectual property rights.*
*Data (e.g. databases) and accompanying explanations will in any case be filed for the purpose of verifying the research. They will be made available for further scientific research in accordance with the guidelines of the University of Groningen's Data Policy and the Data Policy of the research institute.*
*The Dean or the director of the research institute may grant a full or partial exemption from this requirement for a certain period if there are compelling reasons to do so*.

## NWO/European Union

For NWO funding in the Social Sciences a so-called datamanagement contract has to be drawn up post award. This contract includes agreements about archiving data and making them public and is intended to guarantee the quality as well as the availability of the data in the long term. Funding is not paid by NWO until a data contract is signed. Also, a few questions about the data storage and data re-use are processed in the proposal form. Further guidelines (notes) for researchers can be found in the specific calls and are included in the application form.

The EU requires an appropriate consortium agreement to manage (amongst other things) the ownership and access to key knowledge (like data). Further guidelines and regulation are provided in the call.

## Regulations, laws and code

All researchers have to comply to the Dutch law. Several laws are in place, like

the License Law,
the Database Law,
the CAO-NU Intellectual Property Rights and
the General Data Protection Regulation (GDPR), the Algemene Verordening Gegevensbescherming (AVG)

The Netherlands Code of Conduct for Research Integrity translates the law into rules and is therefore a basis for this RDMP.

In case one or more international partners/researchers are involved in the research/project, the laws of the main applicant's residence are applicable in case of any dispute.

## 3. Points of attention

In line with the UG Policy and also anticipating actions laid down in the GDPR workplan 2021, the following points of attention are formulated for the upcoming years.

### Data storage, sharing, accessibility and retention

In accordance with the Netherlands Code of Conduct for Scientific Integrity and the aforementioned UG policy document data are stored for **at least 10 years**.

Data storage is one of the focal points for the upcoming period. At this moment the options are not sufficient to meet the needs of FEB's researchers. Facilities are needed to store data while working on a project but also for archiving data once the project is finished. Facilities to work and share data with others (outside the UG) are very important. The Centre for Information Technology (CIT) currently offers the Y drive and the Virtual Research Workspace (VRW). The Virtual Research Workspace (VRW) consists of a virtualized desktop environment where researchers can work together and safely perform data analyzes. Within the VRW, collaborations are possible with researchers from the organization itself and/or from other institutions. Unfortunately both have their limitations.
The structure of the Y drive is quite complex, and makes it difficult for folder owners to manage access rights. The VRW option is quite expensive and more importantly only allows the use of open source software for users outside the UG.

Also the accesibility of data and the use of repositories need attention. Several options are available like the Open Source Framework, Mendeley, Dataverse and PURE, but for researchers it is difficult to find their way and make the best choice. Clear guidelines need to be developed.

Last but not least the retention periods of data needs attention as well. Policies regarding retention periods for saving data in research have been developed, but there is no protocol yet for timely removing data. In the event of data with personal information the principle of data minimization must be applied as soon as possible, according to the GDPR. There are no clear guidelines in terms of who, how and when.
We plan to develop such a protocol in cooperation with the DCC and the legal department.

## Monitoring data management plans

Each researcher or group of researchers has to fill in a Personal Data Management Plan provided by the research institute SOM. The so-called RDMP tool was implemented in May 2017 and is available online. Per January 1, 2022 almost 1000 plans have been registered.
The tool automatically signals to the researchers and to IRB when a plan has to be submitted to the IRB for ethical approval and/or in case of privacy issues. All plans are listed in an excel file and also checked on a regular basis by the SOM office to identify possible issues. For example in case researchers mention that they store sensitive data on personal devices or in shared drives, they are contacted.

In the fall/winter of 2018, the first audit of our data management practices and ethics policies took place at the UG level, with positive results. The committee used an audit procedure as general guideline to address several topics like policies, implementation processes, practices and training of (young) researchers.

## Training and education

Training and educating researchers and students, with a focus on bachelor, master and PhD students, is very important and several actions are in place or under development.

### *Research master and PhD students*

- In 2020 the workshop scientific integrity offered by the graduate school to PhD and research master students has been redesigned. In addition to integrity issues, the workshop also focusses on data management, ethics and privacy. The focus is on discussing dilemmas.
- To the manual for PhD students a paragraph has been added to inform students about procedures.
- All PhD students receive an exit form around the time of their graduation: one of the items in the questionnaire deals with data, especially where the data are stored when leaving the university. Actions are needed because not all PhD students seem to be compliant with UG regulations.
- To the Teaching and Exam Regulations for the PhD programme an article will be added.
- Currently we are discussing adding a paragraph to the Training & Supervision Plan for PhD students as also suggested in the UG Data Management Polilcy Document.

### *Bachelor and master students*

All master students of FEB have to take the so-called CITI modules as part of their thesis project. The Collaborative Institutional Training Initiative (CITI Programme) of the University of Miami offers eight modules:

- Authorship
- Conflict of interest
- Data management
- Research involving human subjects
- Research misconduct
- Plagiarism

- Privacy/confidentiality
- Internet research

Unfortunately CITI is not GDPR compliant. To get access to the modules students have to register with their personal details and already three years ago we have started the procedure for an agreement together with the legal department. So far the company has refused to sign such an agreement and they refer to their privacy statement.
This issue has recently been discussed with the director of education of FEB. It has been agreed that EU-based alternatives for the CITI modules will be explored and together with the legal department a meeting with the program directors will be organised to discuss possible alternatives.

## Open access of data

Open Science is one of the focal points of the Strategy Evaluation Protocol (SEP). FEB's research is currently evaluated and in the self-evaluation report is laid down that we aim to stimulate 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 organizations to readily gain access to, and use, scientific information in an early phase.
The Peer Review Committee praises the UG for open access policies in general and FEB for results obtained so far. The Committee stated that "there is an excellent open sciency policy at Groningen, with evidence of excellent academic and societal dissemination route". At the same time, the committee emphasizes that we still have a long way to go.

In addition the Board of the UG has recently approved of a large-scale Open Science project. The five pillars of this Open Science Programme are

- Further information of **Open Access** publishing
- Stimulating **FAIR data**, and providing more information on Open Data
- Stimulating the use of **Open Educational Resources**
- Achieve maximum **impact** of science on society by Public Engagement
- Develop an Open Scienc **communication** approach to raise awareness.

The programme will run until 2023 and is ambitious. In line with this programme we will come up with a FEB specific policy document.

## Communication

Communication with researchers and student is very important. We will continue with our activities, like regular newsmail and the manual for students. Other actions are

- Creating a page on infonet with FEB specific information, guidelines (on how to work with data in the broadest sense) and useful links
- Making a leaflet to add to the introduction package for new staff.

## Appendix 1: Definitions

Research/research project
A project conducted by one or more researchers, with the aim to publish the results in any form (including journal publications, dissertations, conference proceedings, working papers, interviews in the media). A research data management plan has to be available at the moment the researchers indeed publish their results. Hence, researchers are required to fill in their individual Data Management Plan (DMP) before that moment.

Raw data
[Field data, Experimental data (field), Survey data:] data including all original variables and observations, recorded in the stage before data cleaning, imputation or any manipulation of the variables. The data may have been edited to put them in an accessible and easy to use format.

[Experimental data (lab):] a file with all decisions made by subjects in an experiment conducted in the FEB Research lab, as automatically registered by the software during the experimental session.

Code
Computer code that describes the entire sequence of steps taken in processing the raw data to the final data set, plus the code that describes how the raw or final data translate into the entries of published tables (e.g. descriptive statistics, regression coefficients etc.), published figures (input to histograms, line graphs etc.) and all published results of statistical tests (p-values, Mann-Whitney test statistics etc.). Examples of such files are STATA do-files, Matlab m-files, and SPSS syntax files.

Log
Text file with the same purpose and content as code. This is for cases where no do-files are available because the data manipulations are done by hand (e.g., for some processing in Excel or SPSS).

Final data
The file obtained from the raw data after the possible cleaning, imputation, and variable manipulation procedures as described in "code" but including all observations present in the raw data. The variables and observations are described in the codebook. The final data file is an easy to access file which still contains all the information used in the analysis on which the publication reports. The final data file is the point of departure for researchers who wish to replicate the findings in the publication or who wish to build on these findings in new original research projects. This is the file suitable for central deposits such as DANS.

Codebook
A summary of the data collection procedure (when was the data collected by whom) plus a complete description of all variables in the final data and the possible values these variables can take on.

Specimen:
[Experiments (lab, field):] A copy of the instructions subjects received by subjects who participated in the experiment (one for each treatment). This can be html-files of the screens, code in an experimental software program (such as z-tree), or written out text of the information

the experimenter gave to subjects. If applicable: a written summary of the procedure on how assistants in doing the experiment have been instructed.

[Field data:] If applicable, a copy of the questionnaire that was administered to respondents.

[Survey data:] A copy of the questionnaire that was administered to respondents.

Readme1st.txt:
Text file (to be opened with notepad, latex or word) with description of:
1. Where the raw and final data, code or log, codebook and specimen are stored.
2. Which software (incl. version) should be used to properly run the code.
3. Contact information of the primary researcher.