Research Data Management Plan GRIP

Introduction

At present GRIP (Groningen Research Institute of Pharmacy) consists of 8 research groups each doing research in a specific pharmacy related research field. The combined research activities of the different groups aim to find new and innovative drugs and therapies and to improve the use of existing drugs. Within GRIP research data needs to be properly archived for purposes of both verification (safeguarding scientific integrity) and safekeeping of valuable datasets. This research data management plan is based on university[1] and FMNS policy[2] and sets out how every scientist within GRIP must deal with research data during the research and once the research project has been completed.

Definition of data

Because of the multidisciplinary nature of research institute collected research data have a wide variety. They can be observational data (f.e. numbers captured in real time) or experimental data (from labs and equipment such as numbers, DNA/ protein sequences, instrumental data files and images (often available in a digital format)), and processed data for example data extracted from large databases such as the IADB (selected, uploaded and stored in text files) or data after data mining or statistical analysis). Data types could include text, numbers, images, audio files, video files, reports, surveys, etc. Each researcher (or research group) 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.

Obviously, data generated by researchers of GRIP need to meet the criteria defined by the Board of the University of Groningen. In concrete terms, data generated by research should have the following

characteristics:

- accurate, complete, reliable, authentic and accompanied by metadata (file describing the data sources in relation to (corresponding sections of) the document);
- securely stored with minimum risk of loss;
- registered in a Current Research Information System (CRIS) such as PURE;
- traceable;
- accessible and citable;
- satisfying legal requirements, criteria for ethically sound research, agreements in partnership agreements and conditions laid down by research funders;
- available for verification and further research once the research is complete and/or the researcher has left the University of Groningen.;
- in principle, a minimum storage period of 10 years.

To ensure that research results are disseminated as widely as possible following the primary publication process, the University of Groningen also has adopted the principle that research data must be made openly available, unless ethical, legal or contractual obligations prevent this.

Verifiability

Verifiability of data is a must. This implies that for published results it has to be made clear:

a) on which data the conclusions are based;
b) if and how data were adapted (and on which reasonable grounds);
c) where and how the data can be verified.

[1] policy document of the Board of the University; RUG Research Databeleid, February 2015
[2] letter of the Dean of FMNS on Research Data Management Plan, 20 October 2014

This means that all those involved in data collection and management will need to meet the standards of good data management and will have to act according to the procedures described in this protocol. This means that each researcher will be responsible for storing his/her data. The most suitable way and format to collect and store the data will be defined together with the PI supervising the researcher, and will be described by each researcher in a personal /project data management plan (PDMP). The chair of each research group will periodically check whether the people within their group do what is agreed upon in this protocol. In future this might be linked to the R&O interview.

In summary:

- Each researcher/research group will need to decide per research project and based on the nature of the research what type of data will be collected and stored, and what file format will be the most appropriate for storage.
- For PhD students, the PDMP should be delivered within 6 months from the start of their doctorate (together with the Introductory Essay) and will be discussed during the 6-month interview.

GRIP Research Data Management Agreements

Responsibilities

> In the context of Data Management the researcher will act according to the act "bescherming persoonsgegevens" and other relevant legal and ethical rules[a].

> The group leader (i.e. the PI) bears the general responsibility for the RDMP in her/his group, and thus has to make sure that master-, PhD-students, postdocs, guest researchers and technicians in their group have defined and follow their PDMP. The group leader is also responsible of adapting the GRIP template to the specific features of the data generated in the group.

> Each individual GRIP-researcher (e.g. master-, PhD-student, postdoc, guest researcher, etc.) who generates data is responsible for implementing his/her data management plan (PDMP) according to the GRIP template and based on what was defined in agreement with their group leader (i.e. the tenured PI). The PDMP should specify a plan for access, use and storage of data at the end of the study.

- GRIP does not dictate how to specifically organize the research unit's data archive nor does it provide preferred formats for data storage or preferred naming of data files. As a result of the different nature of the research within GRIP each tenured staff member has its own responsibility to decide on these details as long as his/her choices are accurate, complete, reliable, and data files are safely stored with minimum risk of loss.

> The chairs of the separate research units ensure that the GRIP-protocol described in this document is strictly followed.

> For externally funded research projects the PI will timely provide sufficient information on the project's data management to the director of the institute. The director will check whether the project's data management plan fits with that of the institute.

> A final decision on the potential deletion of research data is taken by the director of the institute.

Who needs to act

> Attached (Annex I) is a GRIP RDMP-template to be used by each researcher for each separate research project. This protocol is based mainly on the RDMP developed by the 3TUs in the Netherlands[3]. The implementation of the protocol will closely followed and if necessary the protocol and/or template will be adapted.

[3] Based on the template RDMP 3TUs in the Netherlands.
(ttp://datacentrum.3tu.nl/fileadmin/editor_upload/pdf/Data_Management_Plan_.docx)

> Every researcher (Master's student, PhD candidate, postdoc and tenured staff) will fill out a PDMP for each research project he/she is involved in and will do this before the actual start of the project. The PDMP should be updated whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.
> All PDMPs will be created using an online browser-based GRIP RDMP https://rdmp.webhosting.rug.nl. As such the PDMPs are stored at a central location and maintained in a database with support of CIT. In addition:
  - Master students will add the PDMP as an appendix to their master project report[b].
> PDMPs and the associated raw and processed data are available upon request to the chair of the research group, the director of GRIP, FMNS dean and the Board of the University.

What, where and when to store

> All data (raw, preliminary and processed data) underlying an intended publication will be archived and preferably linked to the publication in Current Research Information System (CRIS) of the University of Groningen (PURE). In addition, a file is added containing information on how all these relate to the document. This has to be done in such a way that *a researcher working in the discipline can trace back raw data underlying a publication and understands the origin of figures and tables*.
> Currently, most experimental groups use official hard cover lab journals with a unique number. These lab journals are, where possible, stored in a safe at the research unit, for at least 10 years. Pilots are also run to support the desired transition to electronic lab journals. In the case of using such electronic journals, the software system records the data entry history, allowing additions to be made without manipulation of already stored data.
> Digital raw and processed data underlying a publication is at least stored on the research group's Y-drive. The way in which the research unit's data archive is organized is the responsibility of the chair of the unit. Data should be stored in a 'read-only' format to prevent edit or remove any uploaded file afterwards[c].
> If the amount of data is too large to store the researcher often can only store a selection of the data. In the PDMP he/she will need to discuss argumentations for the selections made.
> Data has to be deposited according to the following guidelines:
  - For publications in peer reviewed scientific journals, conference proceedings, book chapters, patents etc.: The first author should compile a documented archive of all data underlying a publication and store it on the research groups Y-drive within 3 months after the publication appears on paper or online, including publication 'early online'. The supervisor of the first author checks during yearly R&D interviews if the data are stored in accordance with an adequate PDMP.
  - For data collected in the context of a PhD study: The documented data archive of the study should be delivered to the promotor upon handing in the final manuscript for the manuscript committee. The promotor will only sign the approval form of the PhD thesis when the data archive of the study has been handed in. The PhD (or promotor) deposits the data archive of the thesis on the Y-drive within 1 month after the thesis has been handed in.
  - For data collected in the context of an MSc study: All data should be deposited no later than the date of handing in the final version of the MSc-report. A grade will only be awarded for the project [tentamenbewijs] when all data have been provided to the daily supervisor of the project (PhD student, postdoc, staff member etc.). The daily supervisor should deposit the data on the Y-drive no longer than 1 month after the grade has been awarded to the student.
> Data that is collected and stored at an external institute, falls under the responsibility of the external institute and does not need be deposited on the Y-drive. However, this is ONLY the case for raw, primary data. All processed, secondary data such as spreadsheets, databases, scripts, code etc. that is used for the thesis/publication must be saved on the Y-drive of the research group.

- Large primary data sets such as sequencing data that are stored elsewhere in a public database do not need to be deposited on the Y-drive. All secondary data must however be saved and the metadata needs to contain information on where to find the primary data.
- All data, including lab journals, are stored for at least 10 years after publication of the manuscript. In special cases the director has the authority to demand that data is kept beyond the necessary 10-year period.

Data access, sharing and use policy

Every responsible staff member will retain principal legal rights to the data and intellectual property developed under his/her supervision, as long as this is in compliance with University policy. Data access for control purposes will be granted according to the following hierarchical line: staff member, scientific director of the institute, faculty dean, Executive Board of the university. The University will remain co-owner of all saved data, also after a scientist left the University. The institute's scientific director will guarantee access to data belonging to a scientist who has left GRIP. Data that is deposited in public databases are subject to external assessment and subsequently can become available at any time.

Data generated within GRIP of interest to other scientific communities will be made available upon request. When requested, the data will be made available by the leading PI so long as the request does not interfere with the primary publication process, and unless ethical, legal or contractual obligations prevent this.

When access is granted by the responsible tenure track or tenured staff member, data will be made available as soon as it is reasonably possible. Data provision is subject to Dutch law ('auteursrecht') and, if applicable, complies with the non-disclosure agreements of the research project.

Implementation timeline

A first inventory within the institute has shown that there are arrangements in most of the research units on the way data is stored. Agreements however are often only made orally, are not well structured and the execution of the agreements is certainly not regularly checked. The protocol describes a mostly new procedure and time is needed for implementation. It is decided not to start with this protocol retroactively. The following implementation timeline has been agreed upon:

Sept. 16 – sept 17:	Once the browser-based RDMP is ready and the working methods are established, we will do a pilot with RDMP with the PhDs who started from September onwards. If necessary the RDMP and browser-based RDMP will be altered.
January 18:	All staff and visiting researchers of the institute (including PhD student and postdocs) will be required to draw up a PDMP and follow the general procedure as outlined including storage of data
September 18:	the described procedure or an adapted version is implemented for master students
Spring 18:	all other researchers will start with describing a PDMP for each new research project and store it in a file at the secretariat of the research
Early 2018	evaluation of GRIP RDMP procedures

Instruction on the use of the GRIP Personal Data Management Plan

The Personal Data Management Template in Annex 1 has been designed for researchers of GRIP[4]. The template has to be filled for each researcher collaborating in a research project or for research students working on a PhD or Masters project. The Data Management Plan consists of 6 sections. A checklist, providing the most important questions to be answered in your PDMP accompanies each section. The filled plan is then:

- uploaded via the FMNS browser based form (https://rdmp.webhosting.rug.nl (use chrome or firefox only));
- printed and attached to the master project report of degree students.

<u>Summary</u>

In practice, here is what you should do:

⇒ If you are a group leader:

- Make sure that you adapted the GRIP template for the project/personal data management plan (PDMP) to the features of the data generated in your group.
- Make sure that all researchers of your group have defined and implement their PDMP

⇒ If you are a researcher (master-, PhD-student, postdoc, guest researcher):

- Prepare and upload your PDMP within 6 months from your arrival at the University of Groningen
- Store your data in compliance with your PDMP

---

[a] *Researchers working with patient information/materials have additional protocols on privacy related issues to which they must adhere.*

[b] *This seems a feasible option but still has to be discussed and decided by staff responsible for the degree programme in Pharmacy.*

[c] *The institute realises that storage on the Y-drive is not an ideal situation for the longer run since it might be difficult to trace documents (no DOI) and it remains an option to edit or remove any uploaded file. Early 2015 we intent to investigate alternatives and might alter our RDMP accordingly:*

- *we will investigate the way the Y-drive is set up by CIT for the Analytical Biochemistry-group. They use a part of the Y-drive were regular staff can only read data but cannot add, delete or modify data.*
- *we recently learned that PURE might be a potential source for storage of data related to a publication.*

[4] Based on the RDMP 3TUs in the Netherlands, and altered for the research institute GRIP (http://datacentrum.3tu.nl/fileadmin/editor_upload/pdf/Data_Management_Plan_.docx)

ANNEX I

GRIP Personal/Project Data Management Plan dd. March 2017

To be filled by student/researcher and responsible supervisor and thereafter signed and stored according to the agreements within the institute

Title	*Give unique title of the research project*
Start date of project	
End date of project	
1.0 Introduction	*Introduction text to the Research Data Management Plan Groningen research institute of Pharmacy (RDMP*
2.0 Project information	
2.1 Responsible researcher	*Give the name of the responsible researcher for your project Usually this is a scientific staff member of the institute.*
2.2 Research group	*Select the research group in which the research is done*
2.3 Description of the research	*PhDs could use the abstract in the TSP or you could use the abstract of the project proposal*
2.4 Funding body/bodies	*f.e. NWO, UG/Faculty, external scholarship such as CSC or LPDP, etc.*
2.5 Grant number	*Projects financed by external funding bodies often provide a 'grant number'. If your project has such a number please provide it here. If not, leave blank*
2.6 Financial code	*Provide the internal (financial) project code that is allocated to your project. If you have not such a specific code, leave this question blank.*
2.7 Collaborative projects	*Are you participating in a joint project in which another partner (other researcher, research group, institution etc.) is officially involved? Yes/No*
2.7.1 Name collaborator	*If yes, give the name of researcher, name of his/her institute, and work e-mail address. Use bullets in case of multiple partners.*
2.7.2 Responsible for data storage	*Data collected and stored at an external institute, falls under the responsibility of the external institute. If this is (partly) the case in your project, give the name, institute, and work e-mail address of the person that is responsible for data management of that data.*
3.0 Integrity, IP and NDA	
3.1 Academic Integrity	*At the University of Groningen research should be conducted according to the Dutch code on scientific integrity. Select 'Yes' if you are familiar with the Code of Conduct and the Regulations. If you are not familiar with the Code read the document carefully via https://www.rug.nl/about-us/organization/rules-and-regulations/algemeen/gedragscodes-nederlandse-universiteiten/code-wetenschapsbeoefening-14-en.pdf . In case you have any questions interpretation do not hesitate to ask your supervisor*
3.2 Intellectual property rights	*The Intellectual Property (IP) for this research project is owned by the University of Groningen. If there is another agreement you should specify this here. Notice that the ownership of research data must be clarified prior to, or at the beginning of a project since future storage and re-use of research data are directly affected by the intellectual property rights.*
3.3 Non-disclosure Agreement	*Does a Non Disclosure Agreement (NDA) apply to (parts of) the project?*

	A non-disclosure agreement (NDA), also known as a confidentiality agreement (CA), confidential disclosure agreement (CDA), proprietary information agreement (PIA), or secrecy agreement (SA), is a legal contract between at least two parties that outlines confidential material, knowledge, or information that the parties wish to share with one another for certain purposes, but wish to restrict access to or by third parties. It is a contract through which the parties agree not to disclose information covered by the agreement. An NDA creates a confidential relationship between the parties to protect any type of confidential and proprietary information or trade secrets. As such, an NDA protects non-public business information.
4.0 Notable aspects	
4.1 GMO's?	*Does the research include Genetically Modified Organisms (GMO)? Yes/No*
4.1.1 Regulations around GMO's	*Does the research project agree with regulations for genetically modified organisms (GMO) and are GMO materials properly documented and stored?* *Refer to the applicable GMO license(s) for the research group or ask your supervisor, the armico of the group or the biological safety officer in your building for the proper information on the license(s).*
4.2 Animal studies	*Does the research project include studies using animals kept in captivity? Yes/No*
4.3 Human subjects	*Does the project include studies on human subjects? Yes/No*
4.4 Radio isotopes	*Does the project include studies with radioactive chemicals/isotopes ? Yes/No*
5.0 Questions about data	
5.1 Data collection	*Describe the data you will be creating/collecting* *Answer the separate questions below but before you do so check the RDMP policy in your research group. There might be examples available that you can use.* *Checklist:* • *In what file formats will your data be collected?* • *Which tools or software are needed to create/process/visualize the data?* • *What is the estimated size of the data, and what growth rate?*
5.2 Data storage and back-up	*How do you ensure that during your research all research data (raw and processed data) are stored securely and backed-up or copied regularly?* *Answer the separate questions below but before you do so check the RDMP policy in your research group. There might be examples available that you can use.* *Checklist:* • *How will the raw and the processed data be stored and backed up during the research?* • *Which storage medium will you use for your storage and backup? [Network storage? Personal storage media (CDs, DVDs, USBs, portable hard drives)? Cloud storage?]* • *Which is the backup frequency and the number of backups at different locations? Please discuss this with your group leader.*
5.3 Data documentation	*How will your data be documented to help future users to understand and use it?* *Check the RDMP policy in your research group. There might be examples available that you can use. Then answer the questions*

	below. *Explain briefly the folder and file naming conventions that you will use in storing your data. Please use a consistent and logical convention.* *If applicable, indicate what project and/or data identifiers will be assigned (e.g. DOI/ Digital Object Identifier).*
5.4 Data access and sharing	*How will you manage access, security and data sharing after your research has finished?* *Note that the University of Groningen is becoming increasingly convinced that findings from research that has been funded by public money should be made freely available and re-used as much as possible. This applies to both academic publications and research data. Take this in mind when answering the questions below.* *Checklist:* • *Are there limitations on the access of your data? (If so, which? E.g. open/restricted access, embargo period, etc.)?* • *Who controls data access (e.g. researcher, group leader, University, funding institution/company)?* • *If you allow others to use your data, how will the data be shared? In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).* • *How is the privacy of posible test subjects guaranteed? And that of third parties, commercial or other?* • *Any sharing requirements (e.g., funder data sharing policy)?* • *Audience for use? Who will use it now? Who will use it later?* • *If your data have been published or will be published soon make sure that you describe how you can trace back the data* • *Which tools/software are needed to view/visualize/analyze the data?*
6.0 Final remarks and signature	
6.1 Final remarks and signature	*According to university rules data should be stored for 10 years at least. In all cases that you would like to store data that might be used in any publication (including Master and PhD-thesis) longer or shorter than 10 years contact the GRIP scientific coordinator first to ask for approval.* *Thereafter:* • *Check all your answers carefully and discuss them with your supervisor/promotor. When both of you are satisfied with the answers given:* • *Sign the form by providing your name in the following answer box and thereafter* • *'Publish' the Project Data Management Plan (PDMP). Only hereafter you have deposited your PDMP.* • *If you are a master student make a print and add it to your master report.* *Please notice that the PDMP should be renewed whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.*

ANNEX II

Preferred formats for data storage

The strongly preferred way of storing all data 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 other data types, consider using the suggested file formats below (based on the KNAW-DANS Preferred Formats overview, May 2013) for similar reasons of compatibility and future accessibility:

Selecting file formats

All formats of digital files stand the risk of becoming obsolete in the future. If a file format becomes obsolete, it means that the current software will not be able to represent and use the content of the file in the way it was meant to at the time of creation. However, some precautions can be taken. One such measure is to select file formats which have a high chance of remaining usable in the far future.

As a general guideline, DANS considers that the file formats best suited for long-term preservation and accessibility are file formats which are commonly used, which have open specifications, and which are independent of specific software, developers or suppliers. However, it is not always possible to select formats that meet with all of these ideal attributes.

Preferred and acceptable formats

At DANS, we have assessed a number of file formats resulting in a list of preferred formats and acceptable formats. This list will change over time as new formats will be developed and others will fall into disuse. The preferred formats are the file formats which we trust to offer the best longterm guarantees for usability, accessibility and robustness. In principle, DANS expects these formats to be durable for the longer term.

The use of acceptable formats will, for a number of reasons, be allowed in the data archive as well, but long-term preservation of these formats is uncertain. DANS therefore strongly recommends data depositors to deliver their data in the preferred format corresponding to the type of data.

TYPE OF DATA	PREFERRED FORMAT(S)	ACCEPTABLE FORMAT(S)
Text documents	• PDF/A (.pdf)	• OpenDocument Text (.odt) • MS Word (.doc, .docx) • Rich Text File (.rtf) • PDF (.pdf)
Plain text	• Unicode TXT (.txt, ...)	• Non-Unicode TXT (.txt, ...)
Spreadsheets	• PDF/A (.pdf) • Comma Separated Values (.csv)	• OpenDocument Spreadsheet (.ods) • MS Excel (.xls, .xlsx)
Databases	• ANSI SQL (.sql, …) • Comma Separated Values (.csv)	• MS Access (.mdb, .accdb) • dBase III or IV (.dbf)
Statistical data	• SPSS Portable (.por) • SAS transport (.sas) • STATA (.dta)	• R (*)
Pictures (raster)	• JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff)	
Pictures (vector)	• PDF/A (.pdf) • Scalable Vector Graphics (.svg)	• Adobe Illustrator (.ai) • PostScript (.eps) • PDF (.pdf)
Video	• MPEG-2 (.mpg, .mpeg, …) • MPEG-4 H264 (.mp4) • Lossless AVI (.avi) • QuickTime (.mov)	
Audio	• WAVE (.wav) • MP3 AAC (.mp3) (**)	
Computer Aided Design	• AutoCAD DXF version R12 (.dxf)	• AutoCAD other versions (.dwg, .dxf)
Geographical Information	• MapInfo Interchange Fomat (.mif/.mid) • ESRI Shapefiles (.shp and accompanying files)	• MapInfo (.tab and accompanying files) • Geographic Markup Language (.gml)

(*) under investigation

(**) please contact DANS for advice before depositing MP3 audio files