GELIFES - RESEARCH DATA MANAGEMENT

This document[1] describes the **guidelines for research data management** of the Groningen Institute for Evolutionary Life Sciences (GELIFES), in place from 1 January, 2015. The **Research Data Management Plan** (RDMP) is the (obligatory) starting point for all research projects in the institute[2] that are coordinated by a staff member (PI) of the institute. The RDMP should describe the collection, processing, storage and archiving of all project data. A **general RDMP template** based on the default institute regulations is available that must be extended with the administrative details of a project, and should also document any specific arrangements concerning data ownership, storage and archiving that are not conform the institute's standard policy. **Archiving of all (published and unpublished) data** in the institute's data repository is required for all research projects coordinated by an institute staff member according to the **standard data archiving procedure**.

GELIFES RESEARCH DATA MANAGEMENT - SUMMARY

- **Responsibilities** – All researchers within the institute are required to draw up a Research Data Management Plan (RDMP) based on the general template at the start of a research project. The senior staff member (PI) ultimately responsible for a research project also has the final responsibility for correct and timely planning, management and archiving of all project data. The director of the institute is responsible for strict compliance with the institute's data management policy. The data managers are responsible for depositing the data archives in, and maintenance of, the data repository.

- **Timeline** – The obligatory data archiving procedure is in place since summer 2013 for the former CEES groups and since the merger of CEES and CBN, January 2015, for all GELIFES groups. A RDMP will be obligatory **from 1 September 2015** for all new research projects, and for all current research projects with **a planned end date later than 1 March 2016**. Note that projects not requiring a RDMP due to imminent finalization will still require proper archiving of all research data.

- **Data ownership** - The research institute acts as the legal representative of the University of Groningen (UoG) and has control over all research data collected, generated or otherwise acquired within the institute unless different arrangements are clearly documented and approved by the institute director.

- **Data management** - A Research Data Management Plan (RDMP) is obligatory for all research projects (MSc, PhD, and other) and/or subprojects within a larger project (PhD, postdoc) that are coordinated by a staff member (PI) of the institute. The RDMP documents data management of all publications based on the project and of unpublished results that are to be published at a later moment. A standard RDMP template based on the institute's data management policy is available.

- **Data archiving** - All published and unpublished primary and secondary data of a project that is coordinated by a staff member (PI) of the institute are to be archived in the institute repository according to the standard data archiving procedure. Archives will be stored for at least 10 years. Datasets that require open access and persistent identifiers upon publication can be archived in an institute Dataverse[3] repository.

- **Data access** - The purpose of the institute data repository is archive-only; access to archive files is restricted to staff members responsible for the corresponding research projects. The director of the institute, the dean of the FMNS and the Board of the UoG have access to all archives for purposes of verification and audits. Data archives are read-only; additions or corrections are to be uploaded as a new version of the archive.

The GELIFES guidelines for data management are based on the UoG Research Data Policy[4]. The UoG policy incorporates relevant guidelines and regulations from the VSNU Code of Conduct for Scientific Practice (2014)[5], the CAO Nederlandse Universiteiten 2015-2016[6] and the VSNU SEP Guidelines & Requirements 2015-2021[7]. The UoG Research Data Policy is being incorporated in turn in the UoG PhD Regulations (2014)[8] and Groningen Graduate School of Science PhD guide (2015)[9]. In concrete terms, data generated by research should have the following characteristics:

- accurate, complete, reliable, authentic and accompanied by metadata;
- securely stored with minimum risk of loss;
- registered in a Current Research Information System (CRIS) i.e., Pure[10];
- traceable;
- accessible and citable;
- satisfying legal requirements, criteria for ethically sound research [11], 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 departed;
- in principle, a minimum storage period of 10 years.

To facilitate secure storage and easy archiving, GELIFES is currently participating in a joint FMNS/UMCG pilot evaluating the electronic lab notebook eLabJournal[12]. eLabJournal is a Laboratory Information Management System (LIMS) that provides cloud-based processing and storage of data (including proper version management and remote backup) with strict authorization filters on who can access the data per project.

The current UoG research data policy on openness is based on the principle of "Open, unless": following publication, research data should be openly accessible unless conflicting ethical, legal or contractual obligations are in place. In practice, the UoG Board has suggested to follow the opt-out possibilities of the Horizon2020 Pilot Action on Open Research Data[13] on the aspect of openness, i.e., total or partial opting out of the Pilot at any stage of a project for a variety of reasons, namely:

- if participation in the Pilot on Open Research Data is incompatible with the Horizon 2020 obligation to protect results if they can reasonably be expected to be commercially or industrially exploited;
- if participation in the Pilot on Open Research Data is incompatible with the need for confidentiality in connection with security issues;
- if participation in the Pilot on Open Research Data is incompatible with existing rules concerning the protection of personal data;
- if participation in the Pilot on Open Research Data would jeopardise the achievement of the main aim of the action;
- if the project will not generate/collect any research data;
- if there are other legitimate reasons to not take part in the Pilot

---

[4] See http://www.rug.nl/research/search/research-data-office/policy/documents/2015-research-data-policy.pdf
[5] See http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2014).pdf
[6] See http://www.vsnu.nl/files/documenten/CAO/caoNU_2015-2016_English-per_1_July_2015.pdf
[7] See http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf
[8] See http://www.rug.nl/education/phd-programme/promotieregeling/promotiereglement-2014.pdf
[9] See http://www.rug.nl/research/gradschool-science/phdguide/guide2015onlineprospectives.pdf
[10] See http://project.etalage.ub.rug.nl/use-pure/
[11] See http://www.rug.nl/about-us/where-do-we-stand/quality-works/research-ethics
[12] See https://www.elabjournal.com/
[13] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

To comply with the UoG policy on openness, and to accommodate requirements of open access of both funding agencies, such as NWO[14], and scientific journals (e.g. PloS ONE), GELIFES will set up an institute Dataverse repository[15] in addition to the current institute repository. The Dataverse repository is specifically meant for (smaller) open access datasets or subsets that require a persistent identifier upon publication.

ADMINISTRATION & MONITORING

Within the institute, a RDMP is obligatory for all research projects (MSc, PhD, and other) and/or subprojects within a larger project (PhD, postdoc) that are coordinated by a staff member (PI) of the institute. The RDMP should document data management of all publications based on the project and of unpublished results that are to be published at a later moment when the temporary staff member involved may have left to another job. All published and unpublished primary and secondary data of the project are to be archived in the institute repository according to the standard data archiving procedure.

Central administration of RDMPs, monitoring of data archive deposits and maintenance of the repository is coordinated by the institute's office and data managers. Administration of the RDMPs is to be implemented through the local secretaries that routinely keep track of arrival and departure of MSc and PhD students, and/or the local data manager; (copies of) all RDMPs will be stored centrally in the institute's data repository until the web-based UoG option is operational (eta spring 2016). The data managers are Joke Bakker (institute coordinator and repository admin; former CEES groups Microbial Ecology, BESO, Theoretical Biology, Evolutionary Genetics); Jan van den Burg (former CEES groups Animal Ecology, Cocon, MarBEE, MarEcon, Plant Ecophysiology; Leon Steijvers (former CBN groups Behavioural Biology, Behavioural Physiology, Chronobiology, Molecular Neurobiology, Neuroendocrinology).

The senior staff member (PI) ultimately responsible for a research project also has the final responsibility for depositing the RDMP at the start of the project, for ensuring that the RDMP is followed during the project with regard to safe temporary storage and integrity of intermediate research data, and for archiving of all project data in the repository at the end of the project.

For MSc research projects, a grade for the project will only be awarded when all data (including documentation) are deposited with the daily supervisor of the project. For PhD projects, the (first) promotor will only sign the approval form of the PhD thesis when the data archive of the study has been handed in.

The institute's data management regulations are included in the section Academic integrity of the PhD R&D form and the R&D form for scientific and support staff, and are pointed out in the R&D interviews (chapter 3). Application of the data management procedures will be evaluated in the annual R&D interviews. At the institute level the data management procedures will be evaluated in the midterm research reviews.

RESPONSIBILITIES & TIMING

MSc projects

At the start of an MSc research project the student writes a RDMP that is to be added as an appendix to the project report; after approval by the supervisor a copy is sent to the data manager for storage in the central administration. The RDMP should be drafted **no later than two months** after the start of an MSc research project.

At the end of the project all data is archived as indicated in the RDMP, and the archive is sent to the data manager for deposit in the GELIFES repository. All data should be deposited no later than the date of handing in the final

---

[14] See http://www.nwo.nl/documents/nwo/juridisch/regeling-subsidieverlening-nwo
[15] See http://www.rug.nl/research/search/research-data-office/facilities/storage/dataverse-storage

version of the MSc report. The daily supervisor should deposit the data archive **no later than 1 month** after the grade has been awarded to the student.

PhD projects

The first version of the RDMP should be added as an attachment to the Training and Support Plan (TSP) of a starting PhD student **no later than six months** after the start of the project; a copy is sent to the data manager for storage in the central administration. The RDMP should be included in the regular evaluation cycle for PhD students and updated when necessary. In case of major changes in the project planning a new version of the RDMP is required. Updated/new versions of the RDMP are added to then TSP and sent to the data manager for central storage.

The documented data archive of the study should be deposited with the promotor upon handing in the final manuscript for the manuscript committee. Also parts/chapters that have already undergone a previous storage procedure upon publication of the paper are to be included in the PhD study archive as one of the folders for each chapter. The promotor will deposit the data archive of the thesis with the data manager for storage in the repository **within 1 month** after the thesis has been handed in.

Research projects

All staff members (postdocs, assistant, associate, full professors) are obliged to draft a RDMP for each research project **no later than 2 months** after the start of a project. In line with the data stewardship policy of many funding agencies, a RDMP is an obligatory part of the grant application, and will be included in the application that is submitted to the financial coordinator. For all other projects, a copy of the RDMP should be sent to the data manager; for large projects that include several subprojects carried out by different researchers (postdocs, PhDs), a separate RDMP is required for each subproject.

Each MSc, PhD student, postdoc and other researchers (including professors) should compile a documented archive of all primary and secondary data underlying a publication and send this in for archiving in the repository **within 3 months** after the paper appears online, including 'early online' publication.