



Research Data Management Policy

General information

Policy name: Research data management

Policy proposal authors: Hannco Bakker & Rutger Vos

Policy proposal submission date: 26 Oct. 2017

Version: 0.6.x

1. Background and aims

At Naturalis we describe, understand, and explain biodiversity. We do this by providing an open environment to science and by sharing our knowledge as widely as possible. We believe that openness in science is of utmost importance to facilitate the dissemination of knowledge for the advancement of science and society.

Current developments in scientific research practices surrounding data require institutional support. The rise of a societal demand for scientific openness, the pursuit of scientific excellence, and the internal need to use our organization's resources (human power and infrastructure) sensibly and efficiently, raise the need for a well-considered internal policy on data management. In addition, funding agencies such as NWO and EU have introduced increasingly explicit requirements for data management and research data stewardship.

To address these developments, this document establishes our policy on research data management, with the aim to more clearly evaluate and control current and future data storage usage and data management practices. The policy reflects the requirements established by funding agencies (with specific attention to NWO and EU), adheres to the FAIR¹ guiding principles for scientific data management and stewardship, dovetails with the European Open Science Cloud² and its preparations for the GO FAIR initiative, and adopts KNAW's recommendations on responsible research data management and the prevention of scientific misconduct³. The policy also aligns with our institutional mission⁴ and the National Plan for Open Science⁵.

¹ The FAIR Guiding Principles for scientific data management and stewardship

<https://www.nature.com/articles/sdata201618>

² European Open Science Cloud

<https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

³ KNAW's recommendations on responsible research data management:

<https://www.knaw.nl/en/news/publications/responsible-research-data-management-and-the-prevention-of-scientific-misconduct>

⁴ Naturalis mission in Policy Plan 2017-2020:

https://drive.google.com/a/naturalis.nl/file/d/0B30nx0jZoA8_dG9DdDdmSlZUem8/view?usp=sharing

⁵ National Plan for Open Science:

https://www.openscience.nl/binaries/content/assets/subsites-evenementen/open-science/national_plan_open_science_the_netherlands_february_2017_en.pdf

In our institutional policy we strive to fit the principles, recommendations, and requirements promulgated in the broader scientific community into our organization's functioning to an extent that is workable and reasonable. This policy therefore provides guidance and support to Naturalis staff across sectors in the organization, including producers of research output, ICT staff, management, and administrative support. The policy is intended to solve problems, not create new ones. We aim to implement good practices⁶ in research data management as a means to promoting high-quality, reproducible, and reusable research output, while acknowledging and respecting the independence of our researchers, whose careers span multiple institutions and complex networks of collaboration.

2. Terms and definitions

2.1. Compliance

The starting point is to comply with existing regulations, including, but not limited to, intellectual property rights regulations and/or laws, the European General Data Protection Regulation, collective labour agreements (Museum CAO), the Code of Conduct of the VSNU⁷, and the requirements of funding agencies (e.g. NWO, EU, ERC).

2.2. Applicability to Naturalis staff

Naturalis staff with a variety of roles and responsibilities, is involved, whether directly or indirectly, in the production and management of research data. This includes the producers of the data themselves (e.g. research staff, students, interns, lab technicians, analysts, guest researchers, and correspondents), IT and information management staff (e.g. IT infrastructure support, librarians, information management), administrative support staff (e.g. RCO, financial administration, and business development) and senior management.

2.3. Applicability to research data

Research data is defined herein as digital data acquired (captured, generated), processed, and analysed during research⁸. Research data relevant for reproduction, replication, and re-use is defined as all raw data, all data of substantial steps, and the final version of the project⁹. *Research data* also covers metadata and the necessary software, scripts and tools in order to reproduce or replicate the research or study. Not covered in this policy is non-digital data (e.g. field notes, observation forms, etc.). With this policy we strongly support the urge to digitize all field notes, observation forms and other non-digital data.

2.4. Metadata

Metadata is documentation that describes data. Properly describing and documenting research data allows users (including the creator) to understand and track important details of the data. Sufficiently exhaustive metadata also facilitates search and retrieval of data deposited in a repository. Metadata can include both descriptive and administrative content such as sampled species, sampling date and location, instrument and protocol details, file format, creator, license, provenance and version information and much more. Metadata can

⁶ Naturalis has signed the Bouchout Declaration for Open Biodiversity Knowledge Management: <http://www.bouchoutdeclaration.org/declaration/>

⁷ The Netherlands Code of Conduct for Academic Practice [http://www.vsnul.nl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20\(version%202014\).pdf](http://www.vsnul.nl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20(version%202014).pdf)

⁸ The Basic Reproducible Workflow Template: <https://www.practicereproducibleresearch.org/core-chapters/3-basic.html>

⁹ To define data as relevant an example of a decision tree is provided in Appendix 1.

take many different forms, from free text to standardized, structured, machine-readable content. Specific disciplines or repositories may guide or dictate the content and format of metadata, here we strongly support the use of a domain-specific standard.

2.5. Data management

A *data management plan* (DMP) outlines the questions and approaches surrounding data management and, when applicable, within the scope of a funded research project. A DMP could be either a project-based DMP or a personal DMP. Where a project-based DMP is typically drafted and submitted to the funding agency early on in the project and may be periodically evaluated and/or updated, a personal DMP is typically drafted for research data outside of projects. In a DMP, the main types of data, their formats, and their volumes are enumerated. A DMP describes all regulations and practices regarding sharing, archiving and reuse of data. A DMP will consider the best practices as established within the applicable research domains, meet the requirements of the funding agency, and/or comply with the Naturalis DMP templates.

2.6. Data steward

This policy requires the implementation of the role of *data steward*. A data steward has the technical expertise to maintain FAIR data. This includes knowledge about file formats, metadata, version changes therein, and options for converting data from proprietary formats to open standards. The data steward has the function to support staff members in data management, to contribute to the writing of DMPs, and to ensure the continued findability, accessibility, and interoperability of data sets under their purview. The data steward has a controlling function regarding all aspects of data management. The responsibilities of the data steward (as in 4.7) are a collection of tasks that will take place in different locations within the organization, at least within information management (IM), research (O&O) and the research coordination office (RCO).

2.7. Principal Investigator

A *principal investigator* (PI) is a leader of a research project or team. This policy requires the assignment of the role of PI in every new DMP, though in the common case of DMPs for funded research projects this role is already unambiguously assigned during the project proposal drafting process.

2.8. Information Security Officer

An *information security officer* oversees that the internal data storage infrastructure complies with data safety regulations and requirements of this policy and those of funding agencies. When external data repositories are in use, the information security officer assesses whether the external data repositories comply with data safety regulations and requirements of this policy and those of involved funding agencies.

2.9. Data ownership

Whether research data constitutes a form of ‘art’ or a collection of ‘facts’ is construed differently in different jurisdictions, and whether ownership can thus be asserted or transferred at all varies likewise: nobody owns facts, according to some legal opinions. Nevertheless, *data ownership* is defined herein as the capacity to legally assert and/or transfer copyright on data, e.g. in the case of data publication, where copyright may be transferred to a publisher or database, in the context of research collaborations, where copyright may be transferred among collaborative partners, or in the case of employment relationships, where copyright may become subject to employment agreements.

2.10. Data ownership within Naturalis

If the principal investigator (PI) of a project's primary employment is at Naturalis, or Naturalis coordinates the project at an institutional level, the policy outlined here defines this as *data ownership within Naturalis* (whether with the PI or the organization), a situation whose implications are discussed elsewhere in this document.

2.11. Data retention

The *retention period* is the period of time during which research data is managed, stewarded, and stored. The Code of Conduct for Scientific Practice of the VSNU states that the minimum retention period of research data is ten years, which is followed by NWO. However, the ambitions and aspirations outlined in this document expand on this: at Naturalis, the retention period of research data is intended to be indefinite.

2.12. Data publication

Data publication refers to the moment that data is placed either on institutional infrastructure and/or at an external data repository. This in order to publish data as FAIR as possible for external interest as part of the scientific record. *Data publication* includes publication of necessary metadata and minting of globally unique identifiers (GUIDs), e.g. in the form of persistent URLs (PURLs¹⁰) and/or collection-URLs (registration numbers). Data publication may include transfer of data ownership, i.e. copyright.

2.13. Time frame of this policy

This policy represents the ambitions and aspirations of Naturalis for the future and will thus go into effect upon and following their formal acceptance and endorsement throughout the organization. Hence, the policy is not enforced retroactively, but retroactive application of relevant guidelines is encouraged.

3. Guidelines

3.1. Preceding research

3.1.1. Project proposals contain a data management section

A data management section in project proposals briefly describes how data is managed, stewarded, and stored during and after the project. This is distinct from the DMP proper, which is formulated in detail should the project proposal be accepted. In the case of data ownership within Naturalis (as per 2.10), several responsible parties within the organization are committed to the formulation of the data management section. Their respective roles are discussed in section 4. Responsibilities.

3.1.2. Accepted projects require a DMP

A DMP is drawn up at the start of the project before any data is gathered or generated. In the case of data ownership within Naturalis (as per 2.10), various responsible parties within the organization (see section 4. Responsibilities) will facilitate and support drafting of the DMP.

¹⁰ The current approach for collection data is to assign a permanent URL (PURL), which can be resolved through the Naturalis PURL Server: <http://data.biodiversitydata.nl/>

3.1.3. Projects with third parties require a DMP that covers data ownership

If a project is subsidized and/or carried out by extramural parties, the DMP must describe how data management responsibilities are assigned to each participant and define rules surrounding ownership of data, and transfer thereof.

3.2 During research

3.2.1. A DMP is maintained

In the lead-in phase of a research project (typically the first year, if not sooner), a DMP is drafted. If required, this DMP is submitted to the funding agency. In addition, a copy is made available to the relevant parties within the organization for purposes of project management, institutional/infrastructural support, and evaluation and assessment. The DMP is updated as needs arise and the updated version is disseminated accordingly.

3.2.2. Research data is stored

Research (meta)data is stored on appropriate infrastructure. In the case of data ownership within Naturalis (as per 2.10), the ICT department will facilitate and support data storage on internal infrastructure. Storage of research (meta)data only on privately-owned storage media (e.g. USB flash drives, private NAS or RAID setups, portable hard disk drives) is considered insecure and is not accepted. See 3.3.8. for how security is promoted on institutional infrastructure.

3.2.3. Research data is periodically backed up

Research data is periodically backed up as specified in the DMP. In the case of data ownership within Naturalis (as per 2.10), the preferred approaches, i.e. using the institutional infrastructure, make use of the automated back-ups of the Naturalis servers.

3.2.4. Access to research data is enacted through the organizational hierarchy

Line management is responsible for ensuring that their staff provide access to research data. In addition to the hierarchical structure, the researcher may request to provide additional access to his/her data (to e.g. associates) provided this complies with the regulations as described in the DMP.

3.3 After research

3.3.1. Research data is published

Research data is published after a research project is completed, either on institutional infrastructure or at an external data repository that complies with the regulations and requirements in this policy. When published on an external data repository a backup will be placed on the institutional infrastructure. Exceptions will be made for sensitive data, e.g. locations of red list species, patient data, intellectual property, etc. In the case of data ownership within Naturalis (as per 2.10), the ICT department facilitates and supports data storage on institutional infrastructure indefinitely.

3.3.2. Published research data is described by metadata and has GUIDs

Research data is accompanied by sufficient metadata as generated during the project. This metadata complies with appropriate community metadata practices and standards¹¹. Metadata describes aspects of the data such as database record IDs, species names, occurrence or sampling locality, details about lab assays, file formats, etc. The purpose of this metadata is to make data findable (for example, by a search engine that indexes it), accessible (in the sense that the metadata describes how the data can be accessed, for example if it is compressed), interoperable (by clearly recording the file formats that were used), and reusable (for example by helping to interpret the experimental conditions under which the data were recorded). In addition, research data has globally unique identifiers (GUIDs) that allow it to be identified unambiguously. In practice this can be as simple as a URL, which is globally unique by virtue of how the internet works, assuming that this URL is stably maintained and not context-sensitive.

3.3.3. Research data publication follows any applicable embargos

Research data is published post-project subsequent to any embargos as agreed with involved funding agencies and project partners. When no such parties are involved, research data is published after the embargo period specified in the applicable DMP.

3.3.4. Published research data is accessible subsequent to applicable limitations

Access to research data post-project is subsequent to agreements with involved funding agencies and project partners. When no such partners are involved, research data is accessible as per the terms described in the applicable DMP.

3.3.5. Research data publication outside of projects is described in a personal DMP

Research data produced during research with no fixed starting and end point in time is stored and made accessible as described in a personal DMP.

3.3.6. Research data of departing staff members is transferred

Storage and accessibility of research data from leaving staff members, students, guest researchers, and correspondents will follow any applicable agreements between the departing person and Naturalis. Unless agreed otherwise, this involves a transfer of the data to line management of the departing staff member.

3.3.7. Relevant research data is disseminated to collection management

Post-project research data that is relevant to collection specimens (e.g. photos, scans, morphological data, sequence data, field observations, links with publications, etc.) is disseminated to collection management and linked to the CMS (collection management system), provided this complies with all applicable embargos and access limitations.

3.3.8. Data publication follows international guidelines

The platform chosen for data publication follows international conventions and standards, thus requires to guarantee data safety, data integrity and sustainable accessibility to data (e.g. Data Seal of Approval¹², Nestor Seal¹³, Core Trust Seal¹⁴, etc.). Generally speaking, the recommended platforms are those that are most commonly used within a given scientific community and those that promote and implement the FAIR practices to the greatest extent.

¹¹ Refer to Appendix 2 for metadata standards.

¹² Data Seal of Approval: <https://www.datasealofapproval.org/en/>

¹³ Nestor Seal: http://www.langzeitarchivierung.de/Subsites/nestor/DE/Siegel/siegel_node.html

¹⁴ Core Trust Seal: <https://www.coretrustseal.org/>

4. Responsibilities

The following responsibilities apply when ownership of research data falls within Naturalis (2.10). When there are extramural consortium participants responsibilities will be described in the consortium agreements of which the DMP can be a part.

The responsibilities are assigned per staff member, per staff member there is a general order following the subjects: i) Policy and regulations, ii) raising awareness and providing support, iii) provision of facilities and roles, iv) DMP practices and regulations, v) DM practices and regulations and vi) data safety.

4.1. General Director

- The General Director is ultimately responsible for the creation and development of the policy by all involved parties.
- The General Director is ultimately responsible for informing all involved parties of their responsibilities and obligations in order to optimize communication.

4.2. Corporate secretary

- The Corporate Secretary responsible for a periodic review of this policy.
- The Corporate secretary supervises maintaining and updating the research data policy.

4.3. CIO

- The CIO is responsible for the introduction of the role of data steward and has the responsibility that the data steward complies with his responsibilities.
- The CIO is responsible for providing basic facilities for data management and data stewardship practices and regulations (e.g. institutional data repository and other resources).
- The CIO is responsible for alignment of digital data storage and RDM requirements
- The CIO is responsible in cooperation with the financial administration for the planning and budgeting of costs and investments regarding infrastructural components and IT management.

4.4. Managing director O&O

- The managing director O&O is responsible for implementation of these regulations.

4.5. RCO

- RCO is responsible for overseeing that a DMP draft is provided before the start of the project, that an initial DMP is written during the project and that the DMP is regularly updated.
- RCO is responsible in cooperation with the financial administration for budgeting the costs and time investments for data management and data stewardship
- RCO is responsible for providing protocols, manuals and templates for data management and data stewardship practices and regulations.
- RCO is responsible for developing frameworks and foreground and background agreements concerning data implicated in monetization or outreach processes.

4.6. Research group leaders

- Group leaders are responsible for their researchers' compliance with data management regulations.

4.7. Data Steward

- The data steward is responsible for adherence to FAIR data principles.
- The data steward is responsible in cooperation with RCO for supporting and facilitating the writing of a DMP at all stages of the project.
- The data steward is responsible in cooperation with RCO as part of the project dossier for the storage and preservation of all DMPs and substantial versions, which require a minimum preservation of 20 years.
- The data steward is responsible in cooperation with Group leaders and Management O&O for raising awareness of data capacity and storage in order to control the growth in data storage.
- The data steward is responsible in cooperation with Group leaders and Management O&O for raising awareness among researchers of their responsibilities and obligations.
- The data steward is responsible in cooperation with the head of the laboratory for raising awareness among research technicians and students of their responsibilities and obligations.
- The data steward is responsible for dissemination, training, support, and advice on data management practices and regulations.
- The data steward is responsible for providing methods and procedures for controlled access to research data.
- The data steward is responsible in cooperation with the researcher for overseeing that research data is correctly managed, stored and accessible during and/or after the project.
- The data steward is responsible for controlling and providing access to and reuse of data on basis of choices from researchers and ensuring that researchers abide by licenses or terms of use when reusing or sharing data.
- The data steward is responsible for data management and data stewardship (transfer) in the event of staff departures to keep data available.

4.8. Information Security Officer

- The information security officer is responsible for providing, guaranteeing and maintaining data safety in the institutional data repository.

4.9. Researchers

- Researchers are responsible for drafting a DMP before the start of the project, writing an initial DMP during the project and keeping the DMP regularly updated.
- Researchers in the role of PIs are responsible for data management compliance with involved funding agencies.
- Researchers are responsible for data management during and after the project.
- Researchers are responsible in cooperation with the Data Steward for storing relevant and reusable research data (2.3) and making it accessible during and/or after the project. Researchers are aware of data capacity and storage.
- Researchers are responsible for transferring/communicating relevant data for the collection of Naturalis to the collection managers and library.
- Researchers are responsible for making guest researchers, correspondents, students, PhD candidates, and postdocs aware of data management and data stewardship guidelines and policy of Naturalis Biodiversity Center.

4.10. Head of laboratory and research technicians

- Head of laboratory is responsible in cooperation with the Data Steward for managing lab data and making it accessible.

- Research technicians (responsible apparatus manager) are responsible for monitoring usage on internal data storage platforms and liaising with ICT to ensure overflow capacity.

4.11. Collection management and library

- Collection- and library- managers are responsible for the management of research data and metadata coming from or linked to the collection in the CMS (collection management system).

4.12. P&O

- P&O is responsible for overseeing that employment agreements and conditions, should they require additional terms in response to the policy outlined herein, are amended following/according to the relevant legal frameworks.

5. Further reading

- Davies, et al. (2017). Open data and digital morphology:
<http://rspb.royalsocietypublishing.org/content/284/1852/20170194>
- Vos, R.A., Fernandes, P.L. (2017). Open Science, Open Data, Open Source:
<http://osodos.org> doi:10.5281/zenodo.1015288
- Griffin, et al. (2017). Best practice Data Life Cycle approaches for the Life Sciences:
<http://www.biorxiv.org/content/early/2017/07/24/167619>
- Kitzes, J., Turek, D., Deniz, F. (Eds.). (in press). *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. Oakland, CA: University of California Press: <https://www.practicereproducibleresearch.org/>
- Vos, R.A. (2016). Ten simple rules for managing high-throughput nucleotide sequencing data.: <http://www.biorxiv.org/content/early/2016/04/19/049338>
- Berchum, M. van, Grootveld, M. (2016). Het beheren van onderzoeksdata.: https://pure.knaw.nl/portal/files/2929306/IW_IVB475.pdf

Appendices

1. Decision tree for defining the value of data for archiving

Example: "[Selection of data for archiving](#)" by Wageningen University.

Also : "[Selecting research Data](#)" by RDNL

2. Metadata standards

A list of [metadata standards](#) is here provided, however metadata standards are not limited to only this list. This is only an example of possible metadata standards in different fields of study.