Data Management Policy

Institute for Science in Society

Faculty of Science

Radboud University

Established by the ISiS management on June 26th, 2018.

Contents

1. Introduction

Data management refers to the ways in which research data are collected, stored, shared, protected, and made available, either for re-use or verification. The Radboud University highly values clear, accurate, and safe processes of data management. Therefore, it has adopted a general Research Data Management (RDM) policy, that serves as the basis for the detailed data management policies of the separate RU-institutes. A key aim of the policy is that the research data of all RU-publications published in 2020 and beyond, will be stored according to both the FAIR principles and the relevant ethical and legal requirements. FAIR means that data must be Findable, Accessible, Interoperable and Reusable (under specific conditions).[1]

This document describes the ISiS RDM-policy, and provides information about all stages and procedures of data management at ISiS. It aims to fulfil the request of the RU executive board to specify and implement an institutional RDM policy. In various stages of drafting this policy, ISiS researchers were consulted to provide input and comments. A draft version was reviewed by the university's RDM steering group, and by RDM support. This policy is available, alongside the RDM policies of other institutes and faculties, on Radboudnet.

The importance of data management

The three main reasons for a Data Management Policy are (1) privacy laws that protect information related to humans, (2) the enhancement of 'open science' by making datasets available for the scientific community, and (3) enabling the control of the empirical basis of publications. The reliable and secure storage of information related to human participants is particularly important for ethical and legal reasons. As of May 2018, a new European privacy regulation, the General Data Protection Regulation (GDPR), is effective. More information on the GDPR can be found on the RU Privacy & Security website.

Data management at ISiS

Whereas safe storage for data and work in progress is important for all researchers, the variety of research at ISiS requires a differentiated approach to data management. Some of the publications of ISiS researchers are based on texts that are already publicly available, and no further measures are needed

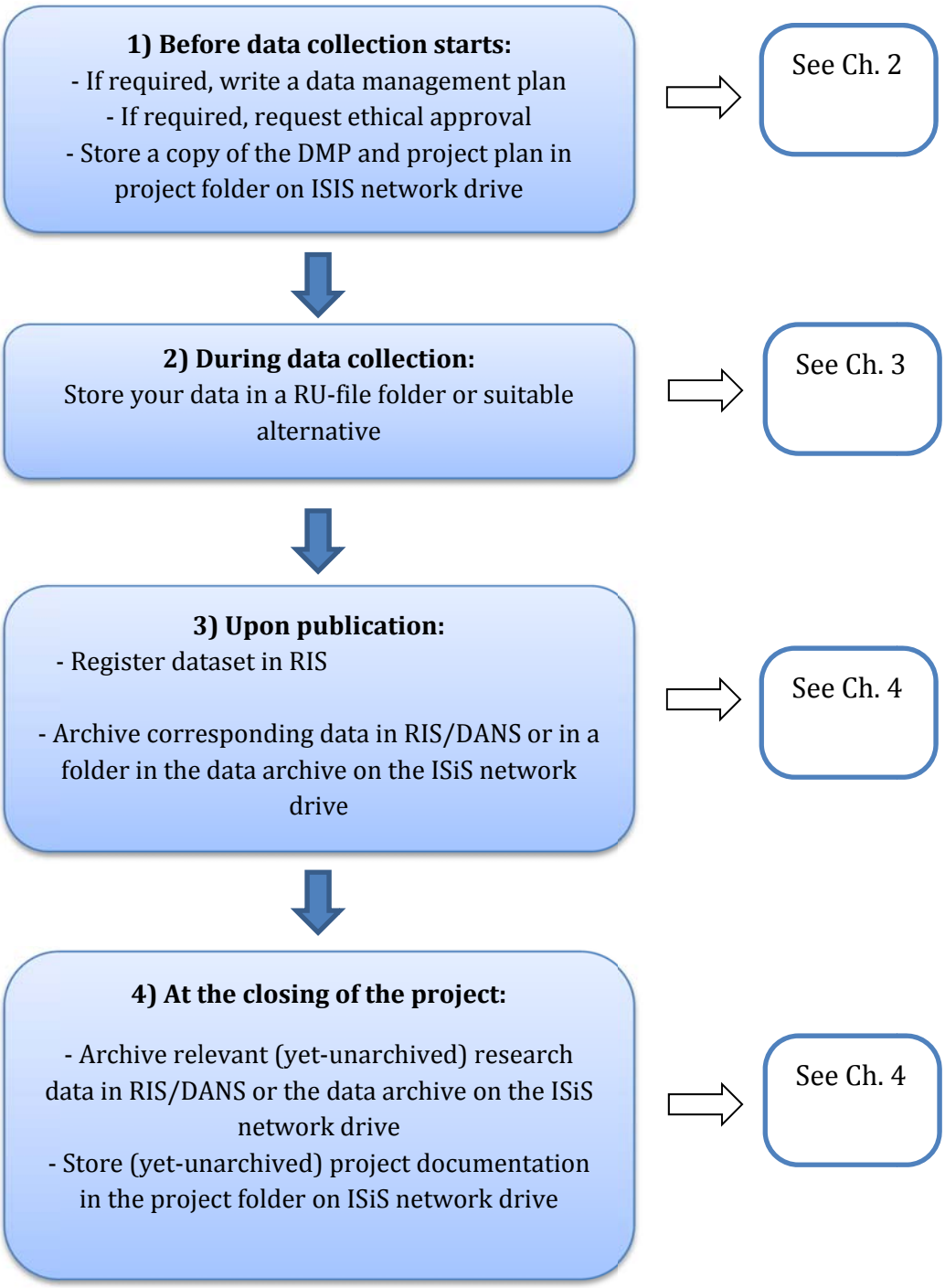
[1] More information on the FAIR principles is provided in Wilkinson et al. (2016) "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data* 3: 160018. doi:10.1038/sdata.2016.18.

to apply the FAIR principles. In other cases, text-based sources can be important to share, and therefore may become subject to data management measures. If publications are based on newly produced empirical data, either about the physical environment or from human participants, further measures for proper storage and sharing possibilities are necessary. Some of these data can be made available to others without limitations, while some cannot be shared publicly, for example due to privacy restrictions or other concerns.

A Data Management Plan (DMP) describes appropriate ways of collecting, storing, sharing and protecting research data. Drafting a DMP is advised for all research that involves not yet publicly available data. However, whether a DMP is obligatory, depends on the type of research. This is further specified in Chapter 2.

The ISiS data steward supports and monitors data management at ISiS. More information about roles and responsibilities is provided in chapter 5. Further support is provided by RDM support (rdmsupport@ubn.ru.nl), who have a website with extensive information on RDM.

Data management steps to be taken by researchers:



2. Data management plan (DMP)

What counts as data?

Data in research may vary between and within the disciplines included in ISiS: documents, observations, measurements, bibliometric data, surveys, interviews, and so on. According to the RU data management policy, it must be clear on what data scientific publications are based. In the remainder of this text the word 'data' refers to those data that are not generally available already (in libraries or in published datasets) and can be made available, where necessary under specified or restricted conditions, at the latest when the research is published.

Personal and sensitive data

Within the Radboud University, a distinction between three types of data is made:
(1) Personal (or critical) data: enables the identification of an individual;
(2) Sensitive data: data that are competition-sensitive or confidential;
(3) Standard data: data that are neither personal nor sensitive.
Directly after collection, personal and sensitive data should be stored in RU-file folders. This is recommended for standard data as well. The request procedure for a RU-file folder is described in chapter 3. More information about personal data can be found on the privacy & security website.

Ethical reflection and/or approval

Ethical aspects arise in all cases in which persons are subjected to specific treatments or rules of behaviour, are asked to give personal and privacy sensitive information, may have to deal with negative consequences of the research, are registered on media like tapes or videos, have to sign a declaration of informed consent, or are minors. Usually in such cases, the Research Ethics Committee (REC) of the Faculty of Science is to ethically assess the research. More information about when ethical assessment is needed, and about the assessment procedure, is provided in the protocol of the REC.

Informed consent

Research that involves human participants in interviews and focus groups, and surveys that include personal data, should in principle use an informed consent (IC) procedure. In such a procedure, participants are informed about the research, how data are managed, and are asked to confirm their

consent, usually in a consent form. The appropriate set-up of an informed consent procedure depends on the method used. For instance, in-depth interviews that involve personal data require a different procedure than online surveys. Both IC procedures and deviations from using such procedures should be approved by the Research Ethics Committee. The website of the REC contains more information on IC procedures.

When to write a data management plan?

In general, researchers should consider their data management in an early stage. A Data Management Plan can be required by a funding organisation or other authority. In other cases, the following principles apply:

- For research that is solely based on publicly available or already published sources, a Data Management Plan (DMP) is not required.
- For research in which standard (non-personal, non-sensitive) empirical data are collected, drafting a DMP is advised, but not obligatory. This may for instance include bibliometric or environmental data.
- For research projects in which personal or sensitive data are collected, a Data Management Plan (DMP) is required. In such cases, the DMP may be part of the information required for the ethical approval.
- For PhD projects that involve any form of empirical data, a DMP is also required. It should be written in the initial stages, before the data collection starts.

The DMP describes how data will be managed, documented, shared, and stored. A DMP can be more or less detailed, depending on the nature and complexity of the research data, and the phase of the research. The DMP helps to make conscious decisions about research data and may save time in later phases of your research. It should be written at the start of research, and can be adjusted during the research process. In case of adjustments, version control is recommended. When written or adjusted, a copy of the DMP should be stored in the project folder on the ISiS network drive. Examples of DMPs and advice in drafting a DMP are available at the ISiS data steward.

How to write a data management plan?

Some funding organisations, such as NWO, have their own obligatory formats for DMPs. For other cases, the Radboud University has a standard format for writing a data management plan. The online DMP tool is useful for drafting a DMP. This tool contains the different DMP formats, and makes it possible for

multiple researchers to write the DMP together and request feedback from the data steward or RDM support.

Chapters 3 and 4 contain additional information that may help drafting a DMP. Examples of DMPs can be obtained via de RIS service desk or the ISiS data steward. If you have questions about writing a data management plan, or need advice or feedback, you can also contact RDM support (rdmsupport@ubn.ru.nl; 024-3611878). There are several training possibilities available.

Important points in the data management plan:

• Data storage during research	Use RU-file folders ('werkgroepmappen') or a suitable alternative.
• Ethical reflection or ethics approval for research concerning human participants	The researcher has to check whether approval from the Research Ethics Committee of the Faculty of Science is required.
• Data archiving upon publication	Data should be stored in DANS, via RIS, in a folder with restricted access in the data archive on the ISiS network drive, or a suitable alternative. Instructions can be found on the RIS-site.
• Anonymity of the participants	Personal data must be stored with care in RU folders with restricted access or suitable alternatives. Such data should in principle be anonymized or pseudonymized, password protected, stored in encrypted location, or deleted, if required.

3. Data storage during collection and analysis

Availability of storage

During your research you can collect different types of research data, such as questionnaire data, experimental data, measurements, observational data, interviews, or texts. This chapter describes where and how the different types of data should be stored. Particularly for research involving personal data, safe and secure storage is crucial; personal data should be accessible only for those with an authorisation. Storage of personal data on personal computers, laptops, tablets, memory sticks, external hard drives and so on, although inevitable in particular cases (such as field work or during an interview), do not meet these requirements, unless your equipment is encrypted. More information on encryption of devices can be found on the privacy & security website.

This RDM policy distinguishes between four main storage facilities:

1. The project archive on the ISiS network drive. Researchers at ISiS are requested to archive essential project documentation in this archive. This includes (to the extent applicable) the project plan or proposal, the DMP, project reports or evaluations, and written output such as articles. This enables the ISiS management to easily find and consult such documents, for instance for self-evaluations. Moreover, it enables the data steward to monitor data management and archiving at ISiS.
2. The RU-file folders (werkgroepmappen), for storing data and sharing work among project team members during the research project (further explained in this chapter). NB: these are not the personal network folders; the latter do not allow for managing access for multiple colleagues.
3. The data archive on the ISiS network drive, for archiving data that cannot be made publicly or conditionally available (chapter 4).
4. Storage in DANS via the RIS interface, for data that can be shared in open or restricted access (chapter 4).

How to request a RU-file folder?

A RU-file folder can be requested at the ICT WebShop. Costs for RU-file folders are covered by the RU. Different types of folders are available: only for employees, or for employees and students. The owner of the folder can manage the access rights of colleagues and/or students. RU-file folders can be accessed

from outside the campus via a VPN connection. For further information, please consult RadboudNet. The organisation of folders is up to the researcher responsible for the folder and may be addressed in the DMP.

How to store your data?

The DMP describes the data types that are stored and the practice and means of storage during research. Clear documentation is important to keep data files understandable for yourself and others in the long run. Such documentation may include code books and separate files that contain metadata describing the data set.

In principle, personal data should be anonymised or pseudonymised as soon as possible in the research process. However, there may be reasons to deviate from this principle. For instance, when experts are interviewed on their topic of expertise, it may be unnecessary or even undesirable to anonymise or pseudonymise interview transcripts or notes.

Pseudonymisation involves using pseudonyms or codes that replace personal information. Fully anonymised or pseudonymised data files do not contain any variables or combinations of variables that can be used to identify individual respondents, such as name, age, address, affiliation, telephone number or BSN (citizen's service number) of respondents. More information can be found on the RDM website. Raw personal data can only be deleted if they are no longer needed for reasons of scientific integrity or reproducibility, or should be deleted if privacy laws or regulations require so.

It may be very difficult to fully anonymise data. For instance, interview transcripts or surveys in which names have been deleted, may still lead to the identification of a respondent. In such cases, the data should be treated as personal throughout the research process and suitable data protection measures should be taken (further specified below).

Where to store data on human participants?

All research data that contain information about participants should be stored safely during the process of collection and analysis. This data should be stored in a RU-file folder ('werkgroepmap') on the university network, or in a suitable alternative (see below). These folders are backed up daily by the ICT department of the RU. For RU-file folders and folders on the ISiS network drive, access rights can be managed per individual.

- Raw data containing personal information (e.g. recordings of interviews or focus groups) should be stored in a RU-file folder (or suitable alternative) that is only accessible to the researcher(s) involved and the data steward. The folder should not be accessible to others.
- In case of pseudonymisation, the document that links the pseudonyms or codes to the personal information (e.g. names of participants) should be stored securely and separately from the pseudonymised data, in the personal drive on the RU-network.
- Data containing indirect information about participants (anonymised data, for example) and metadata about the research (questionnaires, data management plan, early version of the published paper and the like) should be stored in a RU-file folder (or suitable alternative) and, if applicable, can be shared and made accessible for the researchers who are authorised.

Suitable alternatives at Radboud University for safe data storage include the DR (Donders Repository) and DRE (Digital Research Environment). These facilities have advanced features for importing, processing, sharing and archiving data. Another alternative is a folder on the ISiS network drive, to which only the directly involved researchers have access. The data steward and RDM Support can be consulted for more information about data storage facilities.

Collaborations and data sharing with other organisations

In the case of cooperation between universities, it should be agreed to the extent possible that the university where the research data are generated, is entitled to store, manage, and grant access to the data. If research data are generated by two or more partners such that the generation of the data cannot be attributed to one of them, then agreements should be made with the other partners on control regarding the research data in relation to storing, managing, and granting access to the data.

In the case of collaborations with external parties such as companies or NGOs, it is important to consider what data can or cannot be shared with these parties. In the case of personal data, such considerations should be included in the DMP.

Several facilities are available for safely sharing data with colleagues and external parties during the research, including SURFDrive and FileSender. Whether a facility is suitable, depends on the type of data. More information about the options and the recommended security measures can be found on the ICT facilities website.

Storage of non-digital data

During the research, non-digital data, such as filled-in questionnaires and informed consent forms, need to be stored in a secure location at ISiS. This can either be a closet that can be locked, or the central physical storage of the institute.

4. Data archiving

General archiving principles

Good data archiving is important for reasons of re-use, replication, and integrity. At the time of publication, the publication and corresponding data need to be registered via the RIS interface on the RU website. Moreover, the data that correspond to the publication need to be archived, either in RIS/DANS, or in a folder with restricted access on the ISiS network drive. At the closing of the project, both (relevant, yet-unarchived) data and project documentation need to be archived. Data that are already publicly available or that are owned by someone else, need not be registered and archived (again).

The registration and archiving of data consist of the following three steps:

1. **Registration of the publication in Metis**

Academic publications are registered in the information system Metis by employees of the library of science. If a publication needs to be registered quickly, it (or a link or doi) can be sent to infobfac@ubn.ru.nl. Subsequently, the publication will usually be registered within a day.

2. Archiving the data that correspond with the publication in the data archive on the ISiS network drive. The retention period for research data is in principle at least 10 years. This means that raw and processed data that cannot be archived in RIS/DANS, including the documentation and metadata that enable others to understand this data, must be archived in an archive folder on the ISiS network drive for 10 years after publication, unless privacy regulations dictate otherwise. The data steward should have access to these folders.

3. Registering and archiving the data that correspond with the publication in RIS (ris.ru.nl). From the perspective of open science, transparency and the efficient use of resources, it is desirable to make research data publicly available. However, data may be unsuitable for this for several reasons. For instance, data that cannot be fully anonymised or pseudonymised, or that are confidential, competition-sensitive, or owned by others, cannot be shared publicly. Furthermore, the risk of misuse or misinterpretation of data may be a concern. This can be obviated by including rich metadata such as

codebooks or explanatory notes in the archive, or by making the data available under restricted access. In the latter case, someone can file a motivated request for access to the data; subsequently, the researcher decides whether access will be granted. Another way of restricting the (re)use of data is adding a license or data use agreement to (a part of) the data. More information on licenses and data use agreements can be found here. In some cases, a temporal embargo on the dataset may be necessary.

Data and accompanying files that can be made publicly or conditionally available, are preferably archived in in the DANS EASY archive, through RIS. Putting files into the DANS EASY archive is permanent and cannot be changed after deposit. The metadata will always be visible to others. Although RIS/DANS is preferred, other archives can also be used. If the dataset is archived elsewhere, the dataset should still be *registered* using the RIS interface, in order to link it to publications and make it visible in research reports and on your RU profile page. Information about registration and archiving in RIS/DANS is available on the RIS website.

What files should be archived?

If you have stored your files during the project according to Chapter 2, most of the data will already be in the file folder. Upon publication, the corresponding data need to be archived in such a way that their correspondence to the publication is clear.

Data that are not suitable for re-use should be stored for minimally ten years in an archive folder on the ISiS network drive. Data and files available for re-use should be stored for minimally ten years in an accessible archive (preferably RIS/DANS). To guarantee the future readability and usability of the dataset, please use DANS' preferred or acceptable formats. If applicable, the following files need to be archived. Please note that this list is not exhaustive, and the mentioned categories do not necessarily apply to all research projects.

1. **Raw data:** Raw data that contains non-anonymised information about persons should be stored in a folder with restricted access in the data archive on the ISIS network drive, or the physical data archive at ISiS.
2. **Processed data:**
    a. **Anonymised data.** This includes all data that you have collected and anonymised, without further cleaning, aggregating, recoding, composite scores, etc. Fully anonymised data that is suitable for re-use and does not contain (a combination of)

variables that can lead to the identification of participants, qualifies for archiving in RIS/DANS. Check RIS: step-by-step to see which variables should be removed or recoded.

b. **Pseudonymised data.** This includes data in which personal information about participants has been replaced by codes or pseudonyms. Fully pseudonymised data that is suitable for re-use and does not contain (a combination of) variables that can lead to the identification of participants, qualifies for archiving in RIS/DANS. If this data is shared via RIS/DANS, the document that links the codes to the personal information should be stored in a folder with restricted access in the data archive on the ISiS network drive.

c. **Cleaned data file.** The file includes variable names, and clear and complete variable labels and values; without any recoding, composite scores, etc. Corrupt or inaccurate records have been corrected or removed by the researchers.

d. **Analysed data file.** This is the file you used for your analyses in the published paper; including recoding of variables, computed composite scores, etc.

3. **Documentation:**

a. **The questionnaire, survey, topic list and so on used in your published paper and/or a description of the measuring instruments**.

b. **Data processing document used for the analyses presented in the publication.** To reproduce your reported results, a data processing document may be included that describes the steps taken in the analysis. By using the data processing document and the cleaned data file, other researchers should be able to reproduce the analysed data file and reproduce the results presented in the paper.

c. **A codebook with explanations or definitions of the names and values of the variables in your data files.** If you included clear labels and values in your data file, a separate codebook is not necessary.

d. **A description of the data collection process**, for instance a copy of the methodology section of a related article.

e. **A read-me text or documentation file**, which briefly describes the structure of your data and guides other users through your data set (e.g., in which order the files should be opened). It contains an overview of the various folders and data files and a short description of the content of each file.

4. **Published paper.** Please archive a copy of the paper in the project archive on the ISiS network drive. The full-text can also be uploaded in RIS.

Storage of non-digital data

Non-digital data, such as filled-in questionnaires and forms, need to be stored in the secure storage of the institute. Such data need to include documentation about the research project, involved researcher(s), access rights, and retention period

5. Roles and responsibilities

The Research Director of ISiS is ultimately responsible for data management at ISiS. However, project leaders are expected to take responsibility for correct and accurate data management in their project(s). PhD students are responsible for the data they collect for their dissertation, together with their supervisors.

The ISiS data steward is appointed by the ISiS management. The data steward serves as a contact person for RDM in the institute, and monitors and supports data management at ISiS. More specifically, the data steward:

1. guides drafting and/or further developing the research institute's RDM policy;
2. monitors progress in executing the research institute's RDM policy;
3. monitors compliance with the research institute's RDM policy;
4. monitors RDM skills in the research institute, and facilitates training of skills;
5. serves as contact/reference point within the research institute.

Reporting start research

Data management starts at the moment the research plan is made. Writing a data management plan and finding appropriate storage is part of this process. In case data management is required, researchers are requested to inform the data steward about the start of the research project. The data steward supports, if necessary, the writing of a data management plan, and the arrangement of appropriate storage.

Facilities and information

The data steward sees that all facilities needed for data management are available. He/she provides proper information or guides for the researcher to get additional support (RDM Support or otherwise).