Research Data Management Protocol

Institute for Science Education - ISE

- This document discusses the steps that need to be taken for RDM by all (senior and junior) researchers in the ISE.
- ***Very briefly:*** all data should be kept under version control during the research, and the repository archived in the Research Information Services database (RIS) upon publication of research outcomes.
- PhD theses should include a data management paragraph, for more info and a template, refer to https://www.ru.nl/rdm/planning-research/data-management-paragraph/.
- Most important rows in the Table below: 1-5 at the start of a research project, 6-8 during the project, and 9 upon publication of research outcomes.

Phase 1. Creating data

	Topic	Policy	Protocol or practical instructions
1.	Data management paragraph (DM§) and plan (DMP)	The data management plan should ideally be drafted at the start of the research project. In most current funding schemes, a DM§ is now a mandatory part of the research proposal. The DMP details the actual implementation of what was written in the DM§ in terms of infrastructure and procedure. *Explanation*: a data management plan stimulates researchers to make conscious decisions about research data. This prevents problems in a later phase of the research, and saves time over the full research cycle. *Requirements*: if a funder requires so, the format of the funder should be used. Otherwise, the RU format has to be used.	- Practical information and sample text for a DM§. - RU has a DMP format, with elaborate information on each topic of the plan. - The RIS service desk lists requirements of the most important funders and offers various kinds of support, including feedback on DM§'s and DMP's. - PhD students add their DMP as an appendix to their Training and supervisionplan (TSP).
2.	Informed consent form	If data is collected that involves people, an informed consent procedure should be part of the data collection process. *Explanation*: in research involving people, it is an ethical necessity to get participants' informed consent. *Requirements*: informed consent procedures involve informing respondents (written, oral or online) as well as getting their (written, oral	- Most ISE projects do involve the collection of data related to people. Therefore, you need approval by the ethics committee - consult the FNWI ethics committee for formats and sample text. - Further practical information on informed consent and ethics committees at the RU.

		or online) consent. Contact the ethics committee to get approval for your informed consent procedure.	
4.	Training on data management	A basic training session on data management is taken by members of the research staff and students. *Explanation*: a training session increases awareness of and skills in research data management. *Requirements*: training sessions are organized in the context of graduate schools (PhD's), research masters (students) and/or staff meetings (senior researchers).	- Training sessions are initiated by the ISE data management team and lectured by the RIS service desk. - Information can be found on the research institute's intranet (surf folder: \Shared\Science Education\data management and ethics and the RIS website. - For possible training session, click here.
5.	Check for existing data	Researchers should be aware of and, if possible, make use of existing data and at start of the research project. *Explanation*: re-use of existing data is part of scientific research *Requirements*: researchers should know existing repositories in the institute's subject area.	- Check this overview of data repositories by the RIS service desk.
6.	Data storage during research	Ideally, during research, data is kept under version control and stored on CNCZ infrastructure. The use of portable storage solutions (e.g. usb sticks and local computer folders) and cloud services should be kept to a minimum. It is important that data are anonymized or pseudonymised as soon as possible after data collection. Analyses should be conducted on the anonymized/pseudonymized data files. *Explanation*: storage through CNCZ guarantees the secure handling of (critical or sensitive) data and prevents data loss. Data is backed up regularly, stored safely, and can be accessed remotely via a VPN connection. *Requirements*: the choice of storage location depends on the kind of data you collect. In case of critical or sensitive data, options are limited.	- Listed are the various storage and sharing possibilities. Additionally, you can find a decision tree on storage and sharing options. - Practical information on how to protect data away from the confined security of a university campus can be found here. - Information about anonymisation, pseudonymisation and personal data can be found on the RDM website.
7.	Data sharing during research	Ideally, during research, data is shared through the version control system selected in step 6. In specific cases, data can be shared without using version control, in which case a strong preference exists for using the campus network or Surfdrive. In case of critical or sensitive data, data	- Listed are the various storage and sharing possibilities. - Information about anonymisation, pseudonymisation and personal data can be found on the RDM website.

		should always be encrypted before sharing. Also, data files which contain personal data need to be anonymised before they can be shared with others or deposited in a public archive. *Explanation*: sharing by using the campus network or Surfdrive guarantees the secure handling of (critical or sensitive) data. *Requirements*: the choice for sharing tools depends on the kind of data you collect. In the case of critical or sensitive data, options are limited.	

Phase 2. Processing data

8.	Data minimally stored during research & documentation	During research, minimally all data necessary for replication (scientific integrity) is stored *and documented*. *Explanation*: scientific integrity requires researchers to store all data that is needed for replication. Additional documentation is required for the data to be understandable, now and in the future *Requirements*: during research, choices should be made on the structure and documentation of data, including filing, versioning, making backups and documentation of data handling	- Keep the `readme.md` (and ideally a documentation tree) up to date in your data repository - Consider the various types of research data and document the research process - Practical information on structuring and documenting data can be found on the research institute's intranet (surf folder: \Shared\Science Education\data management)

Phase 3. Archiving data

9.	Archiving data RDR	At publication, data should be archived by the researcher in RDR. The RDR can be used for publishing data as well as closed archiving. Simply export a version of your repository and register this snapshot of data in RIS. *Explanation*: making data available for re-use is more and more requested by funders and journals. *Requirements*: RDR is user-friendly, includes support and allows the institute to manage and make available research data. Archiving is	- Why and how: RDR. - Consider the access level to the data; where ISE expresses a preference for *open data* where possible, and *restricted data* as the preferred fall-back option.

		done by contacting the data steward with a request to create a dataset. The data steward will subsequently grant the researcher ownership of the data set.	
10.	Requirements of archiving	Archiving data (what and how) should be done in accordance to the RIS guidelines. *Explanation*: proper data archiving (for the long term) implies data completeness, anonymisation if necessary (e.g. to protect the privacy of respondents), metadating the data and documentation. *Requirements*: at the moment of archiving of the data, the researcher should check the step-by-step guide on the RIS website. What data is archived depends on the perspective: scientific integrity (replication) or re-use of data (open science). RDR does allow for the storage of sensitive data, as it allows for the possibility to turn off accessibility.	- Support and checks are provided by the RIS service desk.
11.	The use of another repository	Although RDR are preferred, other repositories can be used. *Explanation*: RDR meet all criteria of a proper data repository, but international collaboration or conventions in the discipline can be reasons to archive data in another (disciplinary) repository. *Requirements*: alternative repositories preferable meet the criteria of a proper data archive.	- The RDM website provides information on choosing a proper archive.
12.	Retention period	Data is stored for ten years minimally. *Explanation*: the ten years minimum of the RU policy, meets (inter)national guidelines on data storage. *Requirements*: some personal data may only be stored for a short time, or as long as necessary in the research project. Consequently, there could be reasons to delete (part of the) data earlier than the given ten years. However, you should carefully consider this since deleting data could also interfere with proving scientific integrity ('destroying evidence').	- Information on retention periods can be found here. - Information on personal data (and how to deal with it, including anonymization) can be found here.

13.	Deletion of the dataset after retention period	There is no maximum retention period unless the typical nature of the data or agreements with respondents require deletion. *Explanation*: in the case of sensitive or critical data, deletion of data after the minimum period may be necessary. *Requirements*: in case a maximum detention period is relevant and the data is archived/going to be archived in RDR, the researcher should communicate this to the RIS service desk/data steward.	- Information on retention periods can be found here.
14.	Registering data in RIS	If RDR is used, data is already registered in the RIS. If another repository is used, the (main) researcher (or: main RU author of the corresponding article) should additionally use RIS to register the metadata of data as output of the institute. *Explanation*: registration of data refers to the process of registering the data as output to the research institute in the RIS. A link between RIS and RDR is also present. The dataset is also presented on the researcher's profile page. *Requirements*: the researcher is responsible of registering the dataset as output in RIS, either by using RDR to archive his/her/their data (then, it is registered as well), or archiving the data elsewhere and only registering it in RIS.	
15.	Archiving non-digital data	Non-digital data should be archived in accordance with the research institute's guidelines. *Explanation*: the RDM policy, including retention periods, is also applicable to non-digital data. *Requirements*: facilities for archiving and managing non-digital data should be offered by the research institute. Generally, non-digital data could be digitized when possible. The hard-copy could be saved in a locked cabinet to which the research office or the data steward has a key. It is important to register these files (data steward or research office), manage access requests and decide on retention (deletion) periods and corresponding roles .	- More information on archiving non-digital data. - Contact the institute's research office or the data steward to discuss how and where non-digital data will be stored in the research institute. - the official regulation requests to keep the hardcopy for 6 months and then destroy it.

16.	Access control on restricted access datasets	Access requests are initially managed by the depositor of the data, i.e. the (main) researcher (or: main RU author of the corresponding article). *Explanation*: data could be archived with various access levels, amongst others restricted access. Then, the depositor or the data is contacted if an access request to the data is made. In RDR, access requests are handled automatically and the depositor receives a request by mail which he/she may grant or deny. *Requirements*: in case the researcher leaves RU, administration rights should be transferred by contacting the data steward or the RIS service desk.	- Access levels to data and the corresponding roles are discussed here. - Information on transferring administration rights to archived data could be given by the institute's data steward or the RIS service desk.
17.8	Data use agreements	A data use agreement should be added to the data in case there are specific demands made by the researcher or the research institute to re-use of archived data (in RIS or a disciplinary repository). *Explanation*: using a data use agreements allows the researcher or research institute to constraint data re-use. *S*haring your data does not necessarily mean that anyone can do everything with data. One way to restrict the use of data is to add a license or a data use agreement to (a part of) the data. *Requirements*: what agreements are made is up to the researcher or the research institution.	- Information on licenses and data use agreements to restrict access to shared research data is presented in this document. - BJZ (Bureau Juridische Zaken) may assist in drafting a data use agreement. Contact the data steward or the RIS service desk for more information on procedures.

Additional remarks

18.	Research data from bachelor and master's students	The planning and storage of (thesis/ReMa) data and/or code of BA/MA students should be in line with the institute's policy. *Explanation*: since awareness of data management is relevant for students as well - the future generation of researchers - (some of) the data management requirements of the research institute may also apply to (thesis/ReMa) data of BA/MA students. *Requirements*: main responsibility for proper data management lies with the supervisor, together with the student. The institute's requirements should be discussed on relevance for students, particularly: should they write a DMP (or maybe, a condensed version), where should they store their final data (in RIS, in the thesis repository, elsewhere)?	- Students should use version control, just like ISE staff.