RESEARCH DATA MANAGEMENT GUIDELINE SCIENCE FACULTY

PURPOSE OF THIS GUIDELINE

This research data guideline sums up the main conditions of the Utrecht University framework for managing research data, but it mainly means to provide practical handlers to manage research data in accordance with these conditions.

All subjects start with the (more or less) formal description of the UU-framework, followed by an enumeration of subject-relevant topics (focus points) and real life practical examples or references to more info (practical).

This guideline is intended to serve in addition to Institute or research-group data policies and can be used as a reference to the existing policy. RDM-support (www.uu.nl/rdm) and ICT-Beta (https://ict.science.uu.nl/index.php/Storage_research_data) invite you to forward any support question you may have in this process.

MAIN SUBJECTS

1. Responsibilities
2. Data management plan
3. Collecting and Storing research data
4. Archiving research data
5. Data documentation
6. Data Security & privacy
7. Data sharing for reuse

Main subjects

The following subjects are considered indispensable when handling research data by the UU research protocol[1]. Suggested additional guidelines can be found in the UU Pick and choose guidelines at the RDM support team site

More information can be found at the RDM-support website: www.uu.nl/rdm

1.   Responsibilities

Institutes formulate their own specifications within these UU-wide conditions. If those responsible feel the need to significantly depart from this protocol, they must submit a substantiated request to this effect to the Faculty Board.

Principal Investigators (PI) are responsible for ensuring that students, researchers and research leaders in their group abide by the institute policy.

The creators of the data: researchers and students and their supervisors have primary responsibility for good stewardship of their research data.

It is recommended that every research group (or institute) appoints a data-contact who:
- serves as a waypoint within the research group (institute) and as a liaison to faculty or RDM-support experts
- can coordinate the creation of the DMP's and can redirect any questions

Practical

The following support groups and services are provided:

- A Faculty contact for the research groups data contact community. Serves as support and to actively inform the community
- Workshops like 'Research Data Management' and 'Ten Reasons to Share your Research Data' (by RDM Support) for data contacts and researchers.
- At request: custom support by RDM Support/ICT-Beta in setting up a DMP or discuss IT and financial challenges
- University (ITS/Research-IT) and Science faculty (ICT-Beta) provide approved data storage solutions.

---

[1] https://www.uu.nl/en/research/research-data-management/university-policy-framework

2. Data management plan

It is the responsibility of each individual researcher (or, in the case of a group of researchers, the Principal Investigator) to draw up a Data Management Plan (DMP) at the start of the research project and to follow up the agreements made in this plan. The DMP specifies the manner in which research data may be collected, stored, preserved, managed, documented and used.

Focus points

    a. Each research project should have an initial data management plan ready within six months from the start of the project

    b. The following is a set of minimum subjects for describing the management of the research data: How is research data collected, stored, preserved, secured, documented and made available for reuse. More specifically, a research data management plan should:
  - describe the necessary means of storage and how these may be covered through project-, institute-, subsidiary-funding or otherwise
  - make arrangements for the continued operational management of the research data at the end of the research project or termination of the employment contract at the institution;

Practical

Visit the RDM guide on data management planning for practical tools and guidance:
https://www.uu.nl/en/research/research-data-management/guides/data-management-planning

3. Collecting and Storing research data

The data are to be stored safely, protected from unauthorized use and loss.

Focus points

Context: data may also refer to public, derived or random data sets. In that case it may suffice to describe the location or condition(s) whereby the data can be reproduced.

    a. Al researchers use one master copy location for their data. The master copy location must be backed up regularly, at least also to one physically distinct location. Preferably the master copy location has automatic backup.

    b. Raw data is stored separately in a write and access protected folder. Preferably the raw data is time stamped or a checksum is calculated and administered, or access is logged to ensure the integrity of the raw data.

    c. Researchers describe their file naming system and folder structure in a separate file.

    d. Researchers clearly indicate versions of (data-) files and have a separate document with an overview of what retained versions entail.

Practical

Visit the RDM guide on storing and preserving data for practical tools and guidance:
https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data

For cooperating with others, cloud solutions such as SURFdrive, Open Science Framework can be used as master copy location. Note that these solutions are not suitable for privacy sensitive data.

4.   Archiving research data

Research data connected with a research project are to be retained upon its conclusion insofar as relevant for the verifiability of the research. The relevancy for archiving should be judged at the discretion of the Principle Investigator.

Focus points

All archived data underlying publications must be available for verification for at least ten years after publication of the results based on the data, counted from the day of publication.
Context: data may also refer to public, derived or random data sets. In that case it may suffice to describe the location or condition(s) whereby the data can be reproduced.

Practical

Visit the RDM guide on storing and preserving data for practical tools and guidance:
https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data

Which data should be preserved, and/or shared? Consider both primary (RAW) data and secondary (processed) data. Archiving all research data beyond the scope of the project can be a (practical and financial) challenge for big data sets. Try to make a realistic assessment of the total amount of data to be archived and contact faculty support about a solution.

4a. Archiving research data: Data Packages

The full set of research data connected with a research project are to be retained upon its conclusion insofar as relevant for the verifiability of the research. In general, the data package holds all information that is needed to understand and interpret the data, its origin, and its processing, either to verify the results in a corresponding publication or for reuse by others.

Practical

The data package for the purpose of archiving for verification contains all of the following (if applicable):
The (location of-) raw data, a script or exact description how the raw data has been turned to, processed data, a script or exact description of the analyses done to come from processed data to results, a codebook, the study level metadata, a sheet with record level metadata, a research group checklist? check file with checkboxes for all content of the package, a readme file explaining the content and coherence of different files in the study etc.

4.b Archiving research data: Data formats

Archived research data are to be stored in a structure that is suitable for long-term preservation and later consultation. The format and software in which research data are created usually depend on how researchers choose to collect and analyze data, often determined by discipline-specific standards and customs.

Focus points

Data formats should ideally have most of the following characteristics: open documentation, support by many software platforms, widespread use, no compression (or lossless compression), no integrated files/programs/scripts, non-proprietary ('open') format, if such options are available for the type of data stored and do not lead to loss of information.

Practical/Example

Open or standard formats - such as OpenDocument Format (ODF), ASCII, tab-delimited format, comma-separated values, XML - as opposed to proprietary ones. Some proprietary formats, such as MS Rich Text Format, MS Excel, SPSS, are widely used and likely to be accessible for a long time. Be aware

that some formats may become obsolete and what needs to be done to prevent this (i.e. virtualizing the entire setup)
More on preferred data formats:
https://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf

5. Data documentation

The research data will be provided with metadata to describe the data with sufficient clarity and to ensure they are findable for further research.

Focus points

  a. Data for sharing is provided with metadata to describe these data. On this basis the data is findable, and can be assessed regarding its reusability.
  b. In general, all steps in the analyses are documented in such a way that it can be easily reproduced by others how the results have been obtained from the raw data. The analyses to come to processed, ready for analyses data and from processed to results can be described or recorded from script separately.
  c. Data is documented and described in such a way that it is clear and understandable to others.

Practical

Visit the RDM guide on data description in practice for practical tools and guidance:
https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data/data-description-in-practice

6. Data Security & privacy

All research data are to be managed in accordance with the university's requirements governing information security, privacy protection and transparency, in this order of importance.

Privacy sensitive data fall under the Dutch personal data protection act ("Wet Bescherming Persoonsgegevens", WBP) which is to be replaced by the European General Data Protection Regulation (GDPR). Data can also fall under the Dutch Medical Treatment Agreement Act ("Wet Geneeskundige Behandelingsovereenkomst", WGBO), or Medical scientific research act ("Wet Medisch Wetenschappelijk Onderzoek, WMO).

Focus points

  - Researchers who work with privacy sensitive data should familiarize themselves with relevant legislation and sign the confidentiality agreement

Practical

  - Security: visit the RDM guide on Policies, codes of conduct and laws for practical tools and guidance: https://www.uu.nl/en/research/research-data-management/guides/policies-codes-of-conduct-and-laws
  - Privacy: visit the RDM guide on Handling personal data for practical tools and guidance: https://www.uu.nl/en/research/research-data-management/guides/handling-personal-data

7. Data sharing for reuse

Archived research data are to be made available for access and reuse at and outside Utrecht University insofar as is reasonably possible and subject to the proper precautionary measures.

Focus points

a. It is not allowed to transfer exclusive rights to the data, for instance to publishers, without keeping the right to reuse and make the data openly available, unless it is required to acquire external funding.
b. The Principle Investigator should see to it that the data underlying publications is made available for reuse (if this is possible regarding IP and privacy issues).
c. A dataset must not be shared in the case where not all rightful claimants have provided their consent, or if copyright infringements have not been solved.

Practical

Visit the RDM guide on publishing and sharing data for practical tools and guidance:

https://www.uu.nl/en/research/research-data-management/guides/publishing-and-sharing-data