

# Guidelines for Data Management Institute for History

## Introduction

The present document is an elaboration of the Leiden University Data Management Regulations for research that takes place within the Leiden Institute of History. It is intended for researchers who are starting a research project within the Institute for History for which a data management plan may be drawn up and / or researchers who are going to submit a proposal to external funders in which data management may play a role. This document is binding for projects that start after December 2019. In cases not covered by this elaboration, the Scientific Director of the Institute for History has the power to make decisions.

## Leiden University Data Management Regulations (RDUL)

As of 19 April 2016, the Executive Board (CvB) has adopted the 'Leiden University Data Management Regulations' that outlines the framework for a university data management policy. The University has instructed the Faculties to determine whether the further elaboration of these Data Management Regulations takes place at faculty level or at the level of the institutes. The Faculty of Humanities has decided that the data management policy within the faculty will be initiated at institute level.

The RDUL defines "data" as NWO does: data is collected, unprocessed data, as well as data generated by analysis. It concerns both digital and non-digital research data. The CvB states that it must be determined at the level of the discipline which data must be stored.

The RDUL stipulates that **all research** at the university that involves collecting or generating data must have a **data management plan** drawn up in advance, detailing the data management policy of the relevant faculty or institute for the research project in question. This data management plan must guarantee that:

- Research data is stored securely during the research, whereby the integrity, availability and confidentiality of the data are guaranteed.
- Research data can be stored in an accessible, comprehensible, reusable and sustainable way after the research, in combination with metadata, documentation and software required for reuse.
- Data storage takes place in accordance with internationally applicable guidelines, preferably digitally (e.g. via a Trusted Digital Repository) and for at least ten years.

The Executive Board is responsible for providing adequate facilities and support to enable responsible data management. The Academic Director of the institute is responsible for the implementation of the policy; at university level there is a data protection officer who supervises compliance with the policy on the security of research data.

The RDUL sets a number of frameworks for the elaboration of the university regulations at a decentralized level. The elaboration must be in line with the standards that are customary in the discipline, and also record:

- How the responsibilities with regard to data management are assigned within the institute;
- What are the relevant data to store, given the type of research within the institute.
- Policy with regard to data arising from BA, MA and RMA education (and theses).
- How integrity, availability and confidentiality of data are guaranteed during the research.
- The maximum retention period of research data and DMPs.

## Data management within the Institute for History

Research within the institute is characterized by great diversity and, in part, by a rapid development in the ways in which data is used. Most of the research at the institute is largely individual in nature, but projects are increasingly working with research data in teams.

Not all research within the Institute for History (LUIH) falls under the RDUL. Two types of research are wholly or partially exempt from the provisions imposed in the RDUL:

1. For some of the researchers, data management will not play a role in research practice, because they neither collect nor generate data. This applies specifically to historical research based on secondary literature, and to qualitative research based on published texts or archival documents that are bibliographically traceable and publicly accessible. A DMP is unnecessary for this type of research.
2. In addition, there is research in which the data is published in appendices. The institute states that research data that is made public as an appendix to a scientific publication only falls under the RDUL during the research. It is unnecessary to save them separately as data even afterwards. However, a DMP is required for this type of research.

For part of the research however data is collected or generated that fully falls under the RDUL. This concerns in any case:

3. research projects in which several researchers work simultaneously with the same data files during the project.
4. research projects in which data is brought together and analyzed in databases: databases generated by research are by definition covered by the RDUL.
5. research projects in which data is collected through interviews, fieldwork or research in non-public archives and recorded in notes, on photographs, via audio recordings or on film; here, both the (raw) notes and the media files produced fall under the RDUL.
6. research projects collecting data from online databases: the results of a search query should be regarded as research generated data if the query is not reproducible with (exactly) the same version of the database used for the original query.

## Research data

It follows from the above that the research data to which the RDUL relates within LUIH can be divided into several categories. The most important of these are:

- Databases with unique data collections arising from fieldwork or archival study, or collected through research in one or more online databases.
- Media files with unique content arising from fieldwork, archival study or interviews, such as images, audio files and video recordings.
- Notes of replicable observations arising from fieldwork or archival study. This may concern descriptions of objects, situations and documents, and copies of non-public archive documents.
- Notes of non-replicable observations, particularly those resulting from interviews. These can be approved or non-approved notes.

The integrity of these four types of data will have to be guaranteed during the investigation and the accessibility and reusability after the investigation. For data files resulting from interviews or from studies in closed archives, confidentiality must also be guaranteed if necessary. Any data management plan for research in which living persons play a role must work out how the General Data Protection

Regulation (GDPR) will be complied with. Support is available at the university ([privacysupport@hum.leidenuniv.nl](mailto:privacysupport@hum.leidenuniv.nl)).

### **External collaboration**

If there are collaborations between researchers from the LUIH with researchers working elsewhere, the following applies:

- When exchanging existing data sets for research purposes, researchers from the LUIH are not directly responsible for the quality of the storage of data that is collected elsewhere, but are required to request written information about how data is collected, and how it will be stored securely, accessible and reusable.
- In project-based collaborations in which new data will be collected, the working method defined in these guidelines applies as a binding guideline, even if the data to be collected during the project is not primarily managed in Leiden.

### **Education: data arising from BA, MA and RMA courses and theses**

Insofar as teaching by researchers and lecturers affiliated with the LUIH involves the use of primary research data, they are bound by the RDUL and the guidelines given in this elaboration, where education leads to a publicly visible result. The following applies:

1. It is the responsibility of the lecturer or thesis supervisor to ensure that, if necessary, a DMP is drawn up by students and that research data is stored in a sustainable manner after the project has ended.
2. It is the responsibility of the study programs to determine the extent to which they allow research projects of students for which a DMP is required within the study program.
3. Lecturers are not obliged to offer courses in which new research data is to be collected that fall under the RDUL.

In practice, projects requiring a DMP will in most cases be MA theses, as these are normally made publicly accessible through the repository. In exceptional cases, this may concern BA theses, BA Seminars or MA Research seminars from which scientific publications arise. It is recommended, when organizing education in which students are required to set up a DMP, to take into account the workload for the teacher and the feasibility for students.

### **Procedure Data Management Plan**

Procedurally, the Institute for History makes a distinction between two scenarios:

- Current, non-project-related research that takes place within the regular appointment. To this end, researchers and their supervisors periodically evaluate the extent to which current research requires a DMP. The following applies:
  - The IvG applies a "no, unless" policy: by far most of this research falls under the approach described above (1).
  - Where this is not the case, a DMP is drawn up for specific, newly to be collected, datasets of substantial size.
  - Non-project-related data files do not need to be publicly archived until the research has been fully completed. It is not necessary to specify a term for this in the DMP, as long as a periodic evaluation is provided.

- Project-related research with a limited duration for which funding has been obtained from external funding bodies. The following applies:
  - When submitting a funding application, it is established whether, and on what grounds, the drawing up of a DMP is necessary, and what resources are required for this.
  - A definitive data management plan is drawn up within one month of the start of the research project. This is done on the basis of the university template Data Management Plan.
  - During the project, the data management plan is managed and if necessary updated by the researcher, but the Academic Director of LUIH always has direct access to the most recent version of the plan. Version numbering is applied when changes are made, and older versions of the DMP are archived.
  - At the end of the project, LUIH is responsible for archiving the data management plan for the period of twenty years as specified in the RDUL.<sup>1</sup>

### **Data management during the project: integrity, availability and confidentiality**

Within the research as it is done at LUIH, there is a crucial distinction for data management during the research process between projects in which several people work on one and the same dataset, and projects in which this is not done (structurally).

- If several researchers are working on the same dataset, a shared work environment must be used that is accessible to all employees of the project from their workplace.
- If this is not the case, the storage at your own workplace via the ULCN account will in principle suffice.

In principle, for a shared work environment, use is made of the facilities and servers that the university offers internally. In principle, no use is made of the services of commercial providers. As a university point of contact for data management, the Center for Digital Scholarship is the first designated partner when it comes to facilities that transcend the standard workplace.

### **Sustainable storage: metadata, software and documentation**

The RDUL states that research data must be stored in combination with "metadata, other documentation and possibly the software that are necessary for potential reuse of the data".

#### *Metadata*

For the data types with which researchers work at LUIH, the following meta-information needs to be included in the dataset:

- The date on which the data was created
- The date on which the data was last edited
- In the case of media files: place where the recording was made and name of the maker.
- In the case of media files: the size (in pixels) and / or the duration.
- In the case of interviews: name of the interviewee (to be anonymised if necessary).
- Any access restrictions on the individual data.

#### *File formats and software*

With a view to possible reuse (RDUL, article 7) it is obvious that files are archived in the file format in which they were used within the project. For the data types with which researchers work at LUIH,

---

<sup>1</sup> As long as there is no university archive for DMPs.

this usually concerns common file formats of which legibility will not be compromised within the period specified in the RDUL.

The DMP specifies the file formats that play a role in the project. For file formats that can be read with well-archived open-source software, or with standard software packages such as Microsoft Office and Adobe Creative Suite, no further explanation is required. For file types that cannot be read with this software, the DMP must indicate how accessibility is guaranteed, or whether the necessary software must be archived.

#### *Other documentation*

In addition to the data, and possibly software, the RDUL requires additional documentation to be stored. LUIH defines this as documentation relating to the origin, research context and usage history of the dataset as a whole. This will in any case concern:

- Title and duration of the research project
- Name (and contact details) of the PI
- Names of researchers who participated in the creation of the dataset
- Description of the content and nature of the data, including technical details
- Description of the origin of the data and of the collection method used
- Bibliographic details of the publications based on the dataset.

#### **Sustainable storage after the project**

At the end of the project, research data must be stored in a Trusted Digital Repository. LUIH, together with NWO and ERC, applies the standard that a Trusted Digital Repository must be provided with a "Data Seal of Approval" or a Nestor Seal or must comply with the RAC standards (ISO 16363). As the university point of contact for data management, the Center for Digital Scholarship manages the overview of repositories that meet these standards.

LUIH ensures that the provisions for responsible data management that have been made available or accessible to researchers by the CvB in the context of RDUL art. 15 meet the diverse needs of the researchers active within the institute, in particular with a view to research data generated from direct funding.