

Selection of data for archiving

Wageningen University²

Source : Research Data Netherlands¹ &

Research data are preserved for different reasons, for example for further analysis, for verification or because of funder or journal requirements. Since 2015 Leiden University has a policy that states that data should be archived for at least ten years. The policy states also that it is up to the research institutes to decide which data should be kept for verification and reuse³. The following guidelines will help you decide which data to preserve for the long term.

There can be different reasons for preserving research data:

- They are valuable: potential value in terms of re-use, national/international standing and quality, originality, size, scale, costs of data production or innovative nature of the research
- They are unique: the data contain non-repeatable observations
- They are important for history, in particular the history of science
- They are important for non-academic purposes, such as for cultural heritage, museums or other presentations

Next to these general considerations, research funder bodies like [The Netherlands Organisation for Scientific Research \(NWO\)](#) are increasingly demanding for research data to be preserved in order to enable re-use. Furthermore the [Netherlands Code of Conduct for Academic Practice \(VSNU\)](#) (pdf) states that raw data must be kept available for a minimum of ten years.

Selecting research data

When you decide to preserve research data based on the reasons given above, there are several things to keep in mind:

- Technical: which data formats, software (standard or tailor-made: research-specific tools), hardware?

¹ Taken literally from <http://www.researchdata.nl/en/activities/data-management/selecting-research-data/>

² <http://www.wageningenur.nl/nl/show/What-is-a-Data-Management-Plan.htm>

³ <http://blogs.library.leiden.edu/researchdata/policy/>

- Metadata: available and sufficient? Technical information, codebooks, information on data structure, contextual information, information on intellectual property rights, links with publications or related data (in a collaboratory e.g.).
- Data: which data from which point of the digital life cycle: raw data, intermediate data, published data?
- Clarity on intellectual property rights, for example copyright, patent and/or database rights, privacy protection?
- Infrastructure available for preserving the data? Either a data archive or an institutional or thematic repository.
- Costs: how are the costs to be covered for selecting, converting, preserving and making the data available?

<http://www.researchdata.nl/en/activities/data-management/selecting-research-data/>

University of Wageningen Data Management Plan :

