Guidance data protocol

Introduction

In April 2015, the Executive Board (CvB) decided on guidelines for research data management (RDM), the handling of research data, at UvA and HvA. The CvB now asks the centres for applied research and research institutes to develop these central guidelines into local policy proposals and instructions (here designated as 'data protocol)'. The aim of the data protocol is to make clear to all those involved what is expected from them.

This paper was written to support the process and gives a brief explanation of the guidelines. The central guidelines introduce two new matters: the data management plan and the data steward. Furthermore, they suggest the following topics for inclusion in a data protocol: roles and responsibilities; rules and agreements concerning storage, access and archiving; support and training; and an appeal procedure.

If you have any questions, please contact the authors. More information on research data management can be found on the RDM website (http://rdm.uva.n/en / http://hva.nl/rdm)

Data management plan

A data management plan (DMP) is a digital document in which the researcher describes what data will be collected during a research project, how the data will be stored and managed during the project and what will happen to the data once the project is finished. The DMP is written at the start of any new research project. There are various models (templates) and checklists available for writing a data management plan. The completed template is the data management plan. During the project, the plan will be checked regularly to see if it is still up to date or if it has to be adjusted.

The central guidelines make the writing of a data management plan mandatory at the outset of any new research project or programme. The following matters can be laid down in a data protocol:
- Who writes the data management plan for a project?
- How is the writing of a data management plan supervised? To whom should the data management plan be submitted? For information only or for approval?
- How is keeping the data management plan up to date supervised?
- Will the data management plans be stored centrally? If so, where? And who will have access?
- Which matters should at least be discussed in a data management plan?
- Which of the matters that are mentioned in a data management plan, can be handled on the level of the institute or the centre?

For more information on data management plans see the RDM website:

On this page you will also find a UvA template for a data management plan with explanatory notes; a HvA template will shortly be added,

Data steward

According to the central guidelines, a data steward is the person officially responsible for the research data management in one or more research projects. This person has the experience and expertise necessary to advise and support researchers, checks at least once a year how things are concerning research data management and reports on this. According to the central guidelines, the research director, the head of a research group, the principal investigator (the leader of the research project) or a professor can fulfil the role of data steward, and can delegate tasks to one or more other persons.

The data protocol specifies the data stewardship in more detail, for example by determining the following matters:

- Who appoints the data steward?
- Who can take on the role of data steward?
- What is the scope of the task: per research project, research group, research centre/institute?
- What are the duties of the data steward?
- What are the powers of the data steward?
- Can the data steward delegate tasks? If so, which tasks and to whom?
- To whom does the data steward report? In what form? How often?
- How is continuity guaranteed if a data steward leaves?

Roles and responsibilities

Persons who (may) play an advisory, supporting, executive or supervisory role in data management are:
a) Faculty dean, research director
b) Research leader, principal investigator
c) Researcher, professor
d) Doctoral researcher
e) Student
f) Data steward
g) Ethics committee

The data protocol describes who has what responsibilities or tasks in handling the research data within the research centre or institute.

Rules and agreements concerning storage

There are numerous ways to lose data: a file is deleted by mistake, a laptop is stolen, the context of the data is unclear, software has become obsolete so that a file can no longer be opened. Therefore it is imperative that all researchers pay attention to the storing, organising and describing of their data. It is also important to lay these things down on the level of the research institute or centre for applied research. For example:

- Where will the data be stored during the research?
- How will data be stored?
- How will the data be protected from loss, theft, misuse, unauthorised access?

- How will be made sure that it is always and at all times clear how the data have been processed?
- How is version management taken care of?
- What is laid down concerning non-digital data and documentation? How will the relation between non-digital and digital data be guaranteed? For example, completed informed-consent forms, but also samples, scale models, etc.

An example of a data storage protocol is the one from Psychology (UvA).
http://labs.psychologie.uva.nl/en/research/data-storage-protocol/data-storage-protocol.html
For more information on storing research data, see the RDM website:
http://rdm.uva.nl/en/looking-after/looking-after.html.

Rules and agreements concerning access

Current research
Data from current research will naturally be accessible to the researchers directly involved, but there are other parties to whom access may be necessary. The data protocol may state:
- Who have access to data from current research? For example, students taking part in the research project, the data steward, and ethics committee, a fellow researcher who is to continue the research if the original researcher were accidentally to drop out.
- How will this access be organised? Who gives access permission?

Finished research
It is important to UvA and HvA that data from finished research are publicly accessible and available for reuse in new research. Once the project is finished, researchers can see to this by storing the data in a data repository which allows them to be published and assigns a persistent identifier, *i.e.* a unique code which refers/links to the dataset (see also the paragraph on archiving).

Datasets containing sensitive data – personal data, company information, information which, if public, may cause harm – cannot be made publicly accessible. These data, however, may be available for new research. In many data repositories it is possible to publish a description of these data, and to make the data themselves accessible on request only.

The data protocol can describe:
- Which data will be made publicly accessible? Which data will not?
- When will the non-sensitive data become publicly accessible? Who takes care of this?
- How will data be made publicly accessible? Does the research institute or centre have a preference for a specific data repository? If so, which?
- Under what licence will the data be made public?
- Who will handle requests for access to non-public data?
- What are the criteria to honour/refuse requests for access to non-public data?

For more information about publishing research data, see the RDM website:
http://rdm.uva.nl/en/sharing/publishing-data/publishing-data.html

Rules and agreements concerning archiving

The central guidelines make the archiving of research data mandatory. Raw research data must be kept for a period of at least 10 years; processed research data (*e.g.* a dataset which is the basis for a publication) must be kept for a period of at least 5 years after publication of the research. Archiving is done in such a way that the data can be consulted as quickly and easily as possible, both by the original researcher and by any other researchers (either internal or external) who wish to use the data for new research.

The data protocol can take care of practical matters, such as:
- What data are being archived? What are the criteria for archiving?
- Where are the data archived and by whom?
- How are the data archived? What are the requirements for the description of the data?
- Who supervises the archiving and the storage periods?
- Who decides if the storage period is extended?
- What are the criteria to decide to extend the storage period?
- What happens to the data if the storage period is not extended?
- What happens to the data of researchers who leave the research institute or centre?
- Will researchers who are no longer affiliated to the research institute or centre be informed of the extension of the storage period or of the destruction of the data collected by them? If so, how, when and by whom?

For more information about archiving research data, see the RDM website:
http://rdm.uva.nl/en/looking-after/archiving/archiving.html

Doctoral students are mentioned specifically in the RDM guidelines (guideline 17): they must deposit the data which accompany their thesis in a discipline specific, national or institutional data repository. The data protocol can give more detailed instructions, for example:
- Who sees to it that a doctoral student deposits his or her data?
- When must the dataset be deposited?
- Does the institute or centre have a preference for a specific data repository? If so, which?
- Should the dataset, if possible, also be made publicly accessible? If so, under what licence?
- Who registers where the dataset has been deposited?
- Should a copy of the dataset be left at the institute or centre? If so, where will it be stored and who supervises this?
- Who handles any requests for access to the data?

Support and training

The data protocol can describe what support, instruction and training will be provided for whom, and by whom. For example:
- (additional) training of researchers,
- training and support of doctoral students,
- training and support of students,
- training of data steward(s), and
- practical support.

Appeal procedure

In order to ensure that in case of conflicts there will be unambiguous action, it is recommended to formulate an appeal procedure. It can determine who decides on what and on the basis of what criteria, if a decision can be appealed, and if so, how and to whom. Such situations as, for example:

* an (external) researcher is not given access to data in the research institute or centre;
* a researcher who is leaving the institute or centre, wishes to take the data with him or her;
* an (ex)researcher objects to extending the storage period of the data;
* an (ex)researcher does not agree with the destruction of the data.

Finally

The data protocol is the result of agreements about how the research data will be handled, which the researchers of a centre for applied research or research institute have come to in cooperation, within the framework of the central RDM guidelines. The points we have drawn attention to in this paper are not intended as an exhaustive survey of matters which must be regulated, but are suggestions only. A data protocol will be a living document: in order to take care that the protocol will not obstruct the research in practice, but will continue to support it, it will be necessary to test or update the protocol regularly.

If you have any questions about the central RDM guidelines, this paper or research data management in general, please contact the authors of this paper. For practical support, for example, the writing of a data management plan or developing RDM training, please contact your information specialist of the UvA/HvA Library via rdm-support@uva.nl.

Early 2016, the authors of this paper intend to organise a network meeting for data stewards in which this paper will be evaluated. They would like hear if any points are missing or indeed superfluous.