Background to the Data Section

This background information is a brief explanation of some data management issues. The information follows the format of the Data Section (DS). With data management you ensure that your data collection can be found, accessed, exchanged, and permanently stored.

1. Collecting and analyzing data

a. Finding existing data collections
Data collections can be found in online catalogues on the internet. The DANS website provides general information about finding data (http://www.dans.knaw.nl/en/content/data-archive/finding-data) and the option to search within the life sciences. Other sites where you can search and find:
- 3TU.Datacentrum (http://datacentrum.3tu.nl/home/) for biomedical technology.
- There are some 200 Dutch biobanks for which the national catalogue of the Biobanking and Biomolecular Research Infrastructure (BBMRI-NL) is available. It also includes a central data management infrastructure which supports large-scale data analysis in genomics and large-scale biobank analyses (www.bbmri.nl & www.nlgenome.nl).
- For image materials there is a national archive in the CTMM TraIT infrastructure for translational studies (Bio-Medical Imaging Archive or BMIA) which allows secure and long-term storage of (radiological) DICOM research images.
- NeuroInformatics.nl is the platform where Dutch neuroinformaticians meet, collaborate, promote results and share data. It is a member of the international INCF Dataspaces (http://www.incf.org/resources/data-space).
- The Dutch-language website http://www.zorggegevens.nl is a guide to data sources on public health care (registrations, surveys, monitors and long-term (cohort) studies).

b. Protecting sensitive data and privacy
If you want to be able to share your collected research data with third parties, you must guarantee the privacy of all individuals and institutions involved.
This can be achieved by taking the following measures:
- Ask permission to use the information from the individuals and/or institutions (informed consent, opt-out). Also take into account research purposes other than those of the original data collection.
- In secondary analysis of research data, make sure that such analysis is only allowed if the purpose is not incompatible with the purpose for which the data were originally collected.
- Use anonymized or pseudonymous data for research analyses.
- Your institute should have a defined security policy.
- See also the Dutch Data Protection Authority at http://www.dutchdpa.nl/Pages/home.aspx and the Code of Conduct for Health Research at http://www.cbpweb.nl/Pages/ged_fmwv_gez_onderzoek.aspx (in Dutch).

c. Linking
Linking personal data (i.e. referring to one and the same person) which have been collected in various studies or registers, is a way to make better use of data collections, made possible if every record in the dataset has a unique identifying code. This may be a pseudonym, a social security number or a combination of variables which enables personal identification with a reasonable measure of certainty (date of birth, sex, postal code).
One of the challenges is the linking or enriching of existing data files in which different standards have been used for capturing the data. Various solutions are available to make data interoperable, for example by translating or converting them.
d. ICT standards
To be able to exchange research data, it is recommended that you choose ICT standards and data infrastructures that are common to your field or are required under legislation. ZonMw is aware that the digital environment in which researchers operate is in full development. The ICT choices you can make are highly dependent on the opportunities that your institute can offer. You should therefore certainly consult your institute’s ICT and/or Data Manager. It is also recommended to obtain the advice of Research Data Netherlands (http://www.researchdata.nl/), one of its collaborating institutes (DANS, 3TU.Datacentrum, SURF), or the Dutch Techcentre for Lifesciences (http://www.dtls.nl/dtl/). They provide services, facilities for ICT, data storage, and regularly consult with the ICT departments of all Dutch research organizations.

2. Delivering data: Project results

a. Descriptive information (metadata)
In order to be able to find, read and interpret the data, you should attach descriptive information (metadata) to your data collection. ZonMw recommends the use of at least the generic Dublin Core Standard metadata schema on the dataset level. Other possibilities include SNOMED CT and the Data Documentation Initiative. You should also pay attention to editing software and context information.

You can further facilitate reuse by also applying one of the richer metadata schemas that are commonly used in your specific field of study. This will enable a more fine-grained description of the data. Metadata not only enhance searchability, they also promote connections to other data collections (for an overview, see http://biosharing.org/standards).

3. Storing data during and after the project
Data storage and back-up during the project is dependent, among other things, on the facilities of your research institute or the external service provider you may employ. Please consult your central ICT department or the experts of SURF (https://www.surf.nl/en).

4. Making data accessible

a. Sustainable storage
ZonMw prefers long-term archiving to be performed in a sustainable manner. International guidelines are available for sustainable data storage. The most simple set of criteria is the one applied by the international Data Seal of Approval. Those criteria and the corresponding seal of approval are dependent on your field of study. They are explained at http://www.trusteddigitalrepository.eu/.

An institute that stores its data according to these guidelines is considered a "trusted digital repository." In the Netherlands this status has been conferred to DANS and 3TU.Datacentrum. There are still hardly any specific care data repositories with the Data Seal of Approval. Researchers may however turn to existing broad repositories (see the list of Seals). Storage at your own institute is also possible, but make sure to take into account the guidelines of the Data Seal of Approval as much as possible. Please also clarify how you will ensure that the data are effectively accessible and reusable after completion of the project.
b. Choice of repositories and archives
It is recommended that you look around for a repository or archive for long-term data storage at an early stage, so you can take into account its requirements (appropriate file formats and metadata) when building the data collection. Also consider the minimum retention period of five years prescribed by the Code of Conduct for Scientific Research.

If you are used to storing your data in international repositories, you should continue to do so, especially if this is required by magazines or co-funders (examples include ArrayExpress for gene-expression experiments and EGA for genotype-phenotype associate and next-generation sequencing experiments from the European Bioinformatics Institute and the sister databases GEO and dbGaP at the US National Institutes of Health).
This usually involves using discipline-specific normative standards (which determine what information should be reported for reuse), such as MIAME, and formatting standards (which determine how the information should be formatted), such as MAGE-TAB.
Then you will not need to also store the data in a (Dutch) trusted digital repository. These international facilities are already optimized for the often large data sets.

c. Terms of use when making data accessible
To be able to share research data with third parties, you should specify precisely which conditions must be met by research groups that want access to your research. In this respect you can think of arrangements about methodology, publications, the period of authorization, data availability, fees, copyright aspects, et cetera. A good data access setup appears to be the DANS model: http://www.dans.knaw.nl/en/content/dans-conditions-use-reuse-deposited-data.

Below is an example of the Data Section form. Please fill in the form online via the link you were given.
‘Enabling Technologies Hotel’-specific Data Section

Applicant Information

Main Applicant’s Name

Institute

Data Section Author’s Name

Data Section Author’s Position

Grant Application Title

Grant Application Number

E-mail Address
**Data Section (DS)**
Relating to the call for project access to Technology Hotels within the programme Enabling Technologies

**Important: Please read the instructions below before creating the Data Section.**

*The Data Section as a data management guide*

The questions in the Data Section (DS) and the background information above are intended as a guide for planning the data management components that you should take into account when preparing your project.

The Data Section corresponds to the stages of your project:
1. Collecting and analyzing data
2. Delivering data: Project results
3. Storing data during and after the project
4. Making data accessible

For substantive information and an overview of DS requirements, see the background document and the PDF version of the Data Section (link).

*Instructions for completing the Data Section*

1. Answer the questions in the DS and, where possible, include sources, authorities, etc. If you do not know the answer to a question (yet), you can indicate this at the end of each paragraph ("Notes on outstanding issues").

2. Because data management at ZonMw is in its pilot stage, we welcome your questions and comments to further improve our procedures and instructions. You can include them under "Questions and comments on Paragraph X" at the end of each paragraph.

3. Once you have submitted your DS, we will return the completed form to you as a PDF file as soon as possible, both by way of confirmation and for your records.

4. Your Data Section has to be submitted before the deadline of the call for project access to Technology Hotels.

Please note: While working on the DS, you can save your answers it at any time. A link to the partially completed form will then be sent to the e-mail address specified. The data will, however, be kept for only 28 days. If you do not submit your DS within this period, the data will be deleted. It is therefore important to be well prepared before you start completing the DS. Please start by reading the entire form and the background document carefully (link).

---

Start completing the Data Section

☐ Yes, having read the instructions above I will now proceed to complete my Data Section.

---

1. Collecting and analyzing data
When you are going to collect new data or reuse existing data, the privacy of individuals or care facilities involved and the information you collect about them must be protected. Also, the process must be documented in a way that enables replication. Permission has to be given for reusing and/or linking to datasets. Collection and analysis should also comply with applicable regulations. This paragraph contains the requirements you must meet if you use data which may, directly or indirectly, be traced back to individual persons.

Additional information and explanations regarding this paragraph can be found in the "Background to the Data Section".

1.1 Reusing and/or linking to data

1.1.1 I will use existing data in my research.

- [ ] Yes (please specify)
- [ ] No

Data file:

---

1.1.2 I have received permission from the data owner(s) to use their data.

- [ ] Yes (please specify)
- [ ] Not yet

Owner(s):

---

1.1.3 I will link to existing data files.

- [ ] Yes (please specify)
- [ ] No

Data files to be linked:

---

1.1.4 I have received permission from the data owner(s) to link to their data.

- [ ] Yes (please specify)
- [ ] No

Owner(s):

---

1.1.5 I know how to make it technically possible to link the data collected in this project to other data.

- [ ] Yes

Identifier:

---

1.1.6 Restrictions apply to the use of the existing data.

- [ ] Yes

Link to privacy regulations or other relevant page:

---

1.2 Collecting new data
1.2.1 I will have to set up a new data collection for my research.
- Yes
- No

1.2.2 I am going to collect data to add to an existing cohort.
- Yes (please specify)
- No

Cohort:

1.2.3 I can estimate the size of the data collection.
- Yes (please specify)
- Not yet

Number of participants:

1.2.4 I have made an arrangement by which I will receive the research data with some form of consent from the participants. I will inform the participants in a way that enables reuse of the research data for other research (by ZonMw).
- Yes (please explain)
- Not yet

Consent will be obtained by:

1.3 Protecting sensitive data

1.3.1 I know which measures to take in order to protect privacy-sensitive data.

- I will have require everyone involved in personal data processing to sign a confidentiality statement
- I will have the project assessed by a Medical Ethical Committee
- I will register the project with the Dutch Data Protection Authority
- I will anonymize privacy-sensitive research data or use pseudonyms
- I will comply with the privacy regulations of the organization with which I am affiliated
- The organization with which I am affiliated has an information security policy
- Other measures (please specify):

1.4 Research data quality

1.4.1 I guarantee that the research data will be of such quality that they can be interpreted and meet quality standards.

- I will document the research process in order to facilitate replication
- I will perform quality checks to ensure completeness, correctness and consistency of the data
- I am bound by a legal basis, guideline or requirement to register/study the data
- Other measures (please specify):

1.5 ICT standards, e-infrastructure
1.5.1 I have selected a standard to be used for recording data.

☐ Yes (please specify)  ☐ Not yet

Standard to be used:

---

1.6 Collaboration with other research organizations

1.6.1 When collecting research data I will collaborate with partners or data suppliers.

☐ Yes  ☐ No

1.6.2 Agreement has been reached about the right to use research data from the project.

☐ Yes (please attach collaboration agreement)  ☐ N/A

Attached collaboration agreement

---

Notes on outstanding issues

---

Questions and comments on Paragraph 1. Collecting and analyzing data

---

2. Delivering data: Project results

In your grant application you describe the intended results of your project. Data are part of the project results. Those data will have to be described and delivered in such a way that they can be found, read, interpreted and, if necessary, linked with a view to future reuse.

Additional information and explanations regarding this paragraph can be found in the "Background to the Data Section".

---

2.1 As far as data are concerned, the following project results are foreseen.

---

2.2 I will provide descriptive information on the data that have been used and/or collected during my research to ensure that they can be read and interpreted in the future.

☐ Yes on the dataset level at least I will apply the generic Dublin Core Standard metadata schema

☐ Yes, and in addition I will apply a metadata schema that is specific to my field of study

Metadata schema specific to my field of study:

---

2.3 I will allow the data to be linked to other research data.

☐ Yes
3. Storing data during and after the project

For the availability of data during and after the project, it is important to know how you will store and make back-ups of the data during the research, who will be responsible for what, and where you will make the data available in the long term. Issues include size, growth, versioning, security, storage term and cost.

Additional information and explanations regarding this paragraph can be found in the "Background to the Data Section".

3.1 Storing data during research

3.1.1 During the project I will have enough storage sites and capacity and I will have a data back-up available.

☐ I will use my organization's standard facilities for storage and back-up of my data
☐ I will use an external service provider for storing back-ups of my data
☐ I will manage the storage and back-up of my research data myself
☐ Other methods (please specify):

3.2 Long-term data archiving

3.2.1 I have selected a repository or archive for post-project archiving/publishing purposes.

☐ Yes (please specify) ☐ Not yet

Repository/Archive:

3.2.2 The archive selected has an international certification.

☐ Yes (please specify) ☐ No ☐ I don't know

Certification name:

3.2.3 I have already decided to apply the recommended storage term of at least 5 years to my data.

☐ Yes (please specify) ☐ Not yet

Storage term:
3.3 Costs

3.3.1 The costs of archiving (including preparation of the data) are covered by this application.

- Yes
- No (please explain)

I will cover them in another way (please specify):

[Blank Line]

Notes on outstanding issues

[Blank Line]

Questions and comments on Paragraph 3. Storing data during and after the project

[Blank Line]

4. Making data accessible

After completion of the project and any embargo period permitted by ZonMw, your data should be available to other research groups. It is important for the data to be easy to find with clear terms of reuse.

Additional information and explanations regarding this paragraph can be found in the "Background to the Data Section".

4.1 Data availability and terms of use

4.1.1 After the research project (or earlier) I will make my data available for verification and follow-up research.

- Yes, directly after the project
- Yes, after an embargo period
- No (please explain)

Embargo period:

[Blank Line]

Explanation:

[Blank Line]

4.1.2 After the project the data will be publicly accessible.

- Yes
- No
4.1.3 I know which conditions I am going to attach to the use of my data.

☐ I will attach conditions to data use for commercial purposes
☐ I will attach conditions relating to data security
☐ I will make arrangement regarding publications, authorship and the period of authorization
☐ A handling fee will be charged for obtaining research data
☐ I will charge other fees
☐ I will conclude a collaboration agreement with one or more research partners
☐ Restrictions will apply to research questions to be answered
☐ I will lay down other terms of use for third parties
☐ Other conditions (please specify): ______________________________________________________________________

4.2 Data and descriptive information

4.2.1 The data collection and the descriptive information that I make available, can easily be found on the internet.

☐ Yes, but the site has not yet been decided  ☐ Yes, and the site has been decided

Internet site: __________________________________________________________________________

Notes on outstanding issues

________________________________________________________________________________________

Questions and comments on Paragraph 4. Making data accessible

________________________________________________________________________________________

Are you sure that you have completed all the questions and want to submit your Data Section now?

☐ Yes, I have completed the questions and will submit my Data Section by clicking the Submit button below.