Instructions for composing a ZonMw data management plan
(English version)
Instructions for composing a ZonMw data management plan

(English Version)

October 2017
ZonMw is The Netherlands Organisation for Health Research and Development

Progress requires research and development. ZonMw funds health research and stimulates use of the knowledge developed to help improve health and healthcare.

ZonMw’s main commissioning organisations are the Ministry of Health, Welfare and Sport and the Netherlands Organisation for Scientific Research.

For further information on this publication or the programme, please contact Margreet Bloemers through e-mail bloemers@zonmw.nl or by telephone +31 (0)70 349 5237.

Author: Margreet Bloemers/ Adee Bodewes
Date: May 2017
Version: 01
Content

1 Introduction ..................................................................................................................................... 9
2 The questionnaire ......................................................................................................................... 10
3 Guidance ....................................................................................................................................... 19
4 Key items ...................................................................................................................................... 26
5 Need help? .................................................................................................................................... 28
1 Introduction

Databases are an important result of ZonMw projects. Research and the underlying data must be verifiable and reusable for future research or to underpin policy and practice.

Researchers need to arrange, at the end of their project, that their data is findable, accessible, interoperable, reusable and permanently stored. ZonMw follows the broad principles of FAIR data; Findable, Accessible, Interoperable and Reusable. In June 2016, the process of data management is launched under the name "Access to Data – Toegang tot Data" within ZonMw.

Researchers need to focus on data management in all stages of the grant procedure, until the end of their project.

The principles of data management are:

- researchers only fill in databases that are findable, accessible, interoperable and reusable
- researchers actively determine whether they can use existing data in their project

Data management in ZonMw procedures

Researchers must provide access to their data collection according to the ZonMw Grant Terms and Conditions (article 20, Data files)

In ZonMw calls for project ideas and/or grants proposals, ‘Access To Data’ (and the required data management) is mentioned as a relevance criterion.

A researcher needs to show if he makes use of existing data, or collects new data. ZonMw gives advice on these plans.

During the development of the grant proposal, the researcher will orientate on data management (through the format of ZonMw for a data management plan (DMP)) and involves expertise on data management from within or outside his / her institution. After the proposal is granted, the researcher makes a DMP.

During the project, the researcher can adjust the DMP. For example, based on the advice of ZonMw, or as new opportunities for data management arise.

At the end of the project, the researchers show that his efforts have led to reusable data (sustainably stored, findable, accessible and interoperable). He does this by providing some key items.

Some tips

- Choosing a metadata standard is important. Choosing and applying a metadata standard early in the research process, takes less time and effort than at the end of the project. (See question 3.2)

- Cooperation agreement with regard to data sharing can be arranged later but in advance is highly recommended as it contributes to better data reuse. (See question 1.7)
2 The questionnaire

FAIR data
By filling out this format, you will be creating a data management plan (DMP). Executing this plan will ensure that your research results, as well as the underlying data, can be verified and reused. To this end, the data must be findable, accessible, interoperable and reusable. These FAIR principles, developed by Force11 have been partially integrated into this DMP.

ZonMw's DMP format
This DMP format is arranged based on the principles of FAIR data. You will work on the DMP during the various phases of your project, which together make up the components of the research data life cycle (collecting and analysing data; storing data during and after the project; publishing data). You can find the questions connected to the DMP in the report Data: Digitale Diamanten (Dutch). The questions in this DMP format are partly intended to inform you about data management, and partly to record specific information regarding your project. Not every section of the DMP can be answered at the start of your project. You can fill out some of the questions later on. The DMP is a dynamic document, which you will complete and if necessary adjust over the course of your project.

Does your project result in FAIR data?
ZonMw will answer this question at the end of your project. You will provide ZonMw with the DOI code of your data collection or dataset, as well as a number of other key items. These key items are made public.

We ask that you to fill out this DMP with the help of the data management expert of your institution or consortium, such as the data manager, ICT expert or research support staff. You will create your DMP on the DCC platform, and will remain accessible there. This DMP format is developed with the advice of several supporting institutions, such as SURF, DANS, DTL and Data4lifesciences. These organisations can provide you of personal advice regarding the contents of your DMP.

Nb This DMP-format has been developed for a broad spectrum of types of research. Hence, some questions may not be relevant for your research.
1. General features of the project and data collection (10 questions)

In this section, you provide contact details of those responsible for your project and for data management. Furthermore, you give a number of basic features of your data collection, which form a starting-off point for planning your approach to data management. These are:

- the origin of the data
- the size of the data collection
- the type of data
- arrangements with collaborative partners

**Note:** for most of these questions, multiple answers are possible.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1) Project leader contact details:</td>
<td>-</td>
</tr>
<tr>
<td>1.2) I have composed my DMP with the assistance of a data management expert. List his or her name, function, organisation/department, phone number and email address.</td>
<td>o The expert is connected to my department or institution *</td>
</tr>
</tbody>
</table>
|                                                                           | o The expert is not connected to my department or institution                    | *
| 1.3) In collecting data for my project, I will do the following:           | □ Use existing data (please specify) *                                          |
|                                                                           | □ Generate new data                                                               | #
|                                                                           | □ Merge different data files (please specify)                                    | #
|                                                                           | □ Add new data to an existing data set (please specify)                          | #
|                                                                           | □ Use an MDS (Minimal Data Set)                                                   | #
| 1.4) In my research, I will use:                                          | o Exclusively quantitative data                                                   *|
|                                                                           | o Exclusively qualitative data                                                    |
|                                                                           | o A combination of quantitative and qualitative data                             |
|                                                                           | o Other (please specify)                                                          |
| 1.5) I will be reusing or combining existing data, and I have the owner’s permission for using or combining their data. | o Yes, I have permission to use the data                                           *|
|                                                                           | o Yes, I have permission to use the data, but I am required to destroy them at the end of the project |
|                                                                           | o No permission is required, since the data are openly accessible                 |
|                                                                           | o No, I will not be reusing or combining existing data                            |
| 1.6) In collecting new data, I will be collaborating with other parties.   | □ Yes, the new data will be (partly) provided by a project partner or supplier    *|
|                                                                           | □ Yes, I will collect the new data in conjunction with other researchers or research groups |
|                                                                           | □ Yes, we have reached agreements on the user rights of the data used in the project |
|                                                                           | □ No                                                                            |
| 1.7) I am a member of a consortium of 2 or more partners. Clear arrangements have been made. | o No, I am not working with 2 or more partners                                    *|

* Single answer  
# Multiple answers
ZonMw – Instructions for composing a ZonMw data management plan
(English version)
October 2017_01

GO TO https://dmponline.dcc.ac.uk/ TO FILL IN THE QUESTIONNAIRE!

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement</th>
<th>Yes, I am a member of a consortium of 2 or more partners, but clear arrangements have not (yet) been made regarding data management and intellectual property (please explain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8) I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=” in the collection and its size in GB/TB</td>
<td>Yes (please specify)</td>
<td>Not yet (please explain)</td>
</tr>
<tr>
<td>1.9) The following end products I will make available for further research and verification (please elaborate briefly)</td>
<td>□ Raw data</td>
<td>□ Several versions of processed data</td>
</tr>
<tr>
<td>1.10) During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)</td>
<td>Yes, I will make use of my institution's standard facilities for storage and backup of my data</td>
<td>Yes, I will make use of an external provider's services for storage and backup of my data</td>
</tr>
</tbody>
</table>

* Single answer
# Multiple answers
GO TO https://dmponline.dcc.ac.uk/ TO FILL IN THE QUESTIONNAIRE!

2. Legislation (including privacy) (4 questions)

This section is aimed at informing you about *legislations* applying to scientific research, and the conditions imposed by these laws on the collection, use, and publication of data. Although you have made arrangements for these in your project plan as well, you need to realise that they also count for data management.

- In the interest of *scientific integrity*, research results must be verifiable. The [Code of Conduct for Scientific Practice (VSNU)](https://www.vsnu.nl/nl/nieuws/doc/bestanden/2017-07-24/Code%20van%20Conduct%20voor%20Wetenschappelijk%20Varderen%28VSNU%29.pdf) and the Standards of Conduct for Scientific Integrity (VSNU and KNAW) mention the need for research *reliability*, a high quality of data collection and storage, proper documentation of every step taken, a *minimum storage term of 10 years* for all raw data, and the archival of raw data.

- If you are doing *research involving human subjects*, a [Medical Research Ethics Committee](https://www.merc.nl/) will in most cases examine your project. Before collecting and/or using the data, you will have to demonstrate how you will acquire permission from your test subjects (informed consent, question 2) and ensure the *protection of personal information*. To learn more about *legislations on information security*, you can contact the [Dutch Data Protection Authority](https://www.orlo.nl/). Your institution must have a *security policy* covering, among other things, the secure storage of data. If you are building a biobank or using material from an existing biobank, please read the [Code of Conduct for Responsible Use of Human Tissue](https://www.embnet.org/embnet/biobank/).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1) I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.</td>
<td>□ No, I will not be doing research involving human subjects; <em>proceed to section 3 (Making data findable)</em>&lt;br&gt;□ Wet Bescherming Persoonsgegevens (Personal Data Protection Act) and the Code of Conduct for Medical Research derived from it; I will register my project with the Dutch Data Protection Authority&lt;br&gt;□ Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)&lt;br&gt;□ Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act); I will submit my project for review by a Medical Research Ethics Committee&lt;br&gt;□ Wet op de Geneeskundige Behandelingsovereenkomst (Medical Treatment Contracts Act)&lt;br&gt;□ Gedragscode Goed gebruik van lichaamsmateriaal (Code of Conduct for Responsible Use of Human Tissue)</td>
</tr>
<tr>
<td>2.2) I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.</td>
<td>□ Yes (please describe the form this consent takes)&lt;br&gt;□ Yes, and this informed consent allows for the reuse of data (note that in the Code of Conduct for Medical Research, ‘reuse’ is also referred to as ‘further use’)</td>
</tr>
<tr>
<td>2.3) I will be doing research involving human subjects, and I will anonymise or pseudonymise any privacy sensitive data.</td>
<td>○ Yes, the data will be anonymised&lt;br&gt;○ Yes, the data will be pseudonymised&lt;br&gt;○ Other (please explain)</td>
</tr>
<tr>
<td>2.4) I will stick to the privacy regulations of my organisation</td>
<td>○ Yes&lt;br&gt;○ My organisation does not have any regulations in this field, so I will personally make sure that the measures</td>
</tr>
</tbody>
</table>
3. Making data findable (3 questions)

The information covered in this section of the DMP is related to the FAIR principle of **Findable** data. You can use it in two different ways:

1. **To find** data collections during the preparation phase of your project (also see section 1, question 3)
2. **To make** a new data collection **findable** at the conclusion of your project (according to FAIR principles)

Data collections can be found in online metadata catalogues or web portals, along with added descriptive information (metadata). The data themselves, however, are not included there. These are stored in an archive or repository (see section 6, Making data reusable). Archives usually have an advanced search engine, which you can use to find reusable data collections.

Within your field of study, you can **increase the profile** of your data collection through a data hub, help desk, user group, or forum.

This section produces the following **key items**:

- **The location** where you have placed information about your data collection (a metadata catalogue, archive or web portal)
- **The DOI code**, a **persistent identifier** making the dataset traceable and citable.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1) The data collection of my project will be findable for subsequent research (note: this is a <strong>key item</strong>, which you should report to ZonMw at the end of your project).</td>
<td>□ Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)</td>
</tr>
<tr>
<td></td>
<td>□ Yes, it can be found through an online (metadata) catalogue or web portal (please specify)</td>
</tr>
<tr>
<td></td>
<td>□ No, I have not yet chosen an archive or catalogue/web portal</td>
</tr>
<tr>
<td>3.2) I will use a metadata standard for the description of my data collection.</td>
<td>o Yes, I will select a metadata standard from the list published by Datacite (please specify)</td>
</tr>
<tr>
<td></td>
<td>o Yes, I will use a metadata standard specific for my field of research (please specify)</td>
</tr>
<tr>
<td></td>
<td>o No, I have not yet chosen a metadata standard</td>
</tr>
<tr>
<td>3.3) I will be using a persistent identifier as a permanent link to my data collection (note: this is a <strong>key item</strong>, which you should report to ZonMw at the conclusion of your project).</td>
<td>o Yes, I will be using the DOI code</td>
</tr>
<tr>
<td></td>
<td>o Yes, in addition to the DOI code I will be using another persistent identifier (please specify)</td>
</tr>
<tr>
<td></td>
<td>o No, I will not be using a persistent identifier (please explain)</td>
</tr>
</tbody>
</table>

* Single answer
# Multiple answers
4. Making data accessible (4 questions)

As a data producer, you must provide access to your data collection according to the ZonMw Grant Terms and Conditions (article 20, Data files) to make your data Accessible according to the FAIR principles.

If you enable open access to your collection, the data, once registered in an archive or repository, will be freely accessible to any interested parties.

Restricted access is also permitted. This means that you, as data producer, attach certain access conditions to the data collection. Interested parties can submit a request for a data set. You, the data producer, will work with a steering committee to decide on these requests.

The link to the collection's terms of use is a key item.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1) Once the project has ended, my data will be accessible for further</td>
<td>o Yes, immediately</td>
</tr>
<tr>
<td>research and verification.</td>
<td>o Yes, after an embargo period (please explain)</td>
</tr>
<tr>
<td></td>
<td>o No (please explain)</td>
</tr>
<tr>
<td>4.2) Once the project has ended, my data collection will be publicly</td>
<td>o Yes, proceed to section 5 (Making data interoperable)</td>
</tr>
<tr>
<td>accessible, without any restrictions (open access).</td>
<td>o No, there will be access restrictions to my data collection (please explain)</td>
</tr>
<tr>
<td>4.3) I have a set of terms of use available to me, which I will use to</td>
<td>o Yes, my institution employs internationally available terms of use</td>
</tr>
<tr>
<td>define the requirements of access to my data collection once the</td>
<td>o Yes, my institution has drafted a set of terms of use with the help of a legal</td>
</tr>
<tr>
<td>project has ended (please provide a link or persistent identifier;</td>
<td>o Not yet, my institution will draft a set of terms of use with the help of a</td>
</tr>
<tr>
<td>also note that this is a key item, which you should report to ZonMw</td>
<td>legal advisor</td>
</tr>
<tr>
<td>at the conclusion of your project).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4) In the terms of use restricting access to my data, I have included</td>
<td>□ Conditions related to data security</td>
</tr>
<tr>
<td>at least the following:</td>
<td>□ Agreements on methodology</td>
</tr>
<tr>
<td></td>
<td>□ Whether or not the data set may be linked with another data set (for reasons</td>
</tr>
<tr>
<td></td>
<td>of privacy)</td>
</tr>
<tr>
<td></td>
<td>□ The sharing of data for commercial purposes, taking into account the provisions</td>
</tr>
<tr>
<td></td>
<td>of state aid law</td>
</tr>
<tr>
<td></td>
<td>□ Collaboration in using the data set, including agreements on publication and</td>
</tr>
<tr>
<td></td>
<td>authorship</td>
</tr>
<tr>
<td></td>
<td>□ The manner in which the data set can be accessed</td>
</tr>
<tr>
<td></td>
<td>□ The permitted period of use of the data set</td>
</tr>
<tr>
<td></td>
<td>□ The reimbursement of costs, for example in obtaining the data</td>
</tr>
<tr>
<td></td>
<td>□ A steering committee, programme committee or project leader will be charged</td>
</tr>
<tr>
<td></td>
<td>with approving data requests</td>
</tr>
</tbody>
</table>
5. Making data interoperable (3 questions)

Data which are Interoperable (according to FAIR principles) can be linked, exchanged and integrated with data from other data collections. This enriches the collection, and provides interesting opportunities for new research questions.

If you wish to link (or exchange, or integrate) different data sets, you must ask yourself whether this is allowed and whether this is possible.

Is it allowed?
In research involving human subjects, the question is whether combining data is permissible in an ethical-legal sense. You have already answered questions of this nature (regarding the permission and protection of personal data) in section 2.

Is it possible?
This involves the question whether it is technically possible to link, exchange or integrate the data. To make sure that it is, you must choose a metadata standard enabling the exchange of data at record levels. This metadata standard is the key item within the sphere of interoperability.

Finally, the quality of the collection determines whether or not the data are interoperable and reusable. Quality (that is, the full and correct conservation of data) is a point of attention throughout the duration of the project.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1) I will select a machine actionable data format, which will allow other researchers and their computers to read my data collection.</td>
<td>o Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>o No (please explain)</td>
</tr>
<tr>
<td>5.2) I will select a metadata standard to allow my data collection to be linked to other collections (note: this is a key item, which you should report to ZonMw at the conclusion of your project).</td>
<td>o Yes, I will select a metadata standard from the list published by Biosharing (please specify)</td>
</tr>
<tr>
<td></td>
<td>o No (please explain)</td>
</tr>
<tr>
<td>5.3) I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.</td>
<td>o Yes, the participants have given their permission for reuse of the data, and the data have been pseudonymised</td>
</tr>
<tr>
<td></td>
<td>o No (please explain)</td>
</tr>
</tbody>
</table>

* Single answer
# Multiple answers

GO TO https://dmponline.dcc.ac.uk/ TO FILL IN THE QUESTIONNAIRE!
6. Making data reusable (7 questions)

The goal of data management is to ensure that data collections are verifiable and Reusable (according to FAIR principles). All sections of this DMP format contribute to this goal. In this section of your DMP, you must show how you will make your data accessible and reusable in the long term. Therefore, you must take proper care of the quality of the collection, document the data collection process, and select the right data for long-term archival and sustained storage. Keep in mind the necessary maintenance of the data collection and updates of the required software in the long term.

### Questions

<table>
<thead>
<tr>
<th>Questions</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1) I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).</td>
<td>□ I will document the research process (please explain)</td>
</tr>
<tr>
<td></td>
<td>□ I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)</td>
</tr>
<tr>
<td></td>
<td>□ In addition, I will take further quality assurance measures (please specify)</td>
</tr>
<tr>
<td></td>
<td>□ I will document the software used in the course of the project (please specify)</td>
</tr>
<tr>
<td>6.2) I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9)</td>
<td>□ Yes</td>
</tr>
<tr>
<td></td>
<td>□ Not yet</td>
</tr>
<tr>
<td></td>
<td>□ Some or all of the data must be destroyed once the project has ended, because of a contract or law</td>
</tr>
<tr>
<td>6.3) Once the project has ended and the data has been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.</td>
<td>□ Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>□ Not yet (please explain)</td>
</tr>
<tr>
<td>6.4) I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)</td>
<td>□ Yes, and this archive has a data seal of approval (please specify the archive)</td>
</tr>
<tr>
<td></td>
<td>□ Yes, and this archive has a different form of certification (please specify the archive and certification)</td>
</tr>
<tr>
<td></td>
<td>□ Yes, and this archive meets certification criteria and intends to get certified (please explain how your data will remain accessible and reusable in the long term)</td>
</tr>
<tr>
<td></td>
<td>□ Not yet</td>
</tr>
<tr>
<td>6.5) Once the project has ended, I will uphold the recommended data preservation period of at least 10 years.</td>
<td>□ Yes, in accordance with VNSU guidelines (please specify the number of years)</td>
</tr>
<tr>
<td></td>
<td>□ Yes, in accordance with other guidelines (please explain, and specify the guidelines and the number of years)</td>
</tr>
<tr>
<td></td>
<td>□ No, specify the guidelines and the number of years (please explain)</td>
</tr>
</tbody>
</table>

* Single answer
# Multiple answers
GO TO https://dmponline.dcc.ac.uk/ TO FILL IN THE QUESTIONNAIRE!

| 6.6) Data management costs during the project and preparations for archival can be included in the project budget. These costs are: | o  Amount ........ (please elaborate)  
|                                                                                                                      | o  Unknown (please explain)          |
| 6.7) The costs of archiving the data set once the project has ended are covered.                                      | o  Yes (please elaborate)            
|                                                                                                                      | o  Not yet (please explain)          |
3 Guidance

We have drawn up guidance notes to help you answer some of the questions. The guidance provides information on the question, and may also include links to websites or organisations that are relevant to data management.

**NB**: guidance is not provided on every question.

<table>
<thead>
<tr>
<th>Question</th>
<th>Guidance</th>
</tr>
</thead>
</table>
| 1.3 | Data collections can be found in online metadata catalogues and data archives. Examples:  
• The [DANS website](#) provides general information on how to find data on health research and life sciences  
• RIVM’s [Zorggegevens.nl is a website with](#) information on data sets on public health and healthcare  
• Statistics Netherlands ([CBS](#)) has lots of [data sets](#) on health and welfare  
• [4TU Datacentrum](#), incl. for biomedical research.  
• Biobanking and Biomolecular Research Infrastructure ([BBMRI-NL](#)) contains 200 Dutch [biobanks](#) ([www.nlgenome.nl](#) & [www.biobanken.nl](#)). BBMRI-NL also provides an infrastructure for data management for large-scale data analysis of genomics data and biobanks.  
• The Netherlands Institute for Sound and Vision for information regarding image libraries.  
• [Bio-Medical Imaging Archive or BMIA](#) is an image library.  
• [NeuroInformatics.nl](#) is a platform where neuroinformatics experts collaborate and share data. |
| 1.4 | The verification and reuse of data is important for both quantitative and qualitative data. Management of these two types of data differs in some respects. The questionnaire draws attention to this fact. |
| 1.5 | When using **existing data** consider: whether conditions apply to the use of the data and whether conflicts could arise with existing research; the fact that for the reproducibility of results you need to note which version of the existing data sets you use.  
Convert or translate different standards in pre-existing data sets in order to link or enhance them. |
| 1.6 | If you are collaborating with other organisations on the gathering of research data, it is important to:  
• make clear arrangements concerning the accessibility, reusability, exchangeability and verifiability of the new data set;  
• agree on **ownership or co-producership** of data;  
• draw up **terms and conditions** for the use of data by third parties (e.g. co-authorship, permission for research, information on new research reports and papers); |
- allocate **responsibilities** in the research process;
- record all these arrangements in writing in the form of a **collaboration or consortium agreement**.

**For more information on ZonMw public-private partnership and co-funding:**

### 1.8

Use "n=" (number of participants or subjects) and the number of giga-/terabytes per participants (or for the entire collection) to estimate **the size of the data collection**. You will need this to plan for things like storage capacity and budget (See part 6 Making data reusable and permanent storage).

### 1.10

Please consider the following when it comes to data storage:

- The properly organised **storage** and **backup** of data is necessary to prevent data from becoming lost due to technical problems or human error.
- Seek advice on **your research institute’s storage facilities**. Contact your IT department, SURF, or any external service provider your organisation uses.
- Storage on laptops, hard disks or external media is generally risky. It is preferable to use robust, properly managed storage facilities provided by the **institution’s IT department**. Automatic backups by the IT department are also safer than manual backups.
- When using **external services** you must ensure that there is no conflict with the policy of the body/ies funding the research (e.g. accessibility of data) or with the policy of your department or institute (e.g. security of sensitive data).

If you **manage** the storage and backup of your research data **yourself**, always consider: the scale of and growth in your data set, storage locations and capacity, version management, backups and technical and organisational measures to secure the data.

### 2. Legislation (incl. privacy)

#### 2.1

The following legislation applies to research on human subjects:

- Personal data include any details relating to an identified or identifiable natural person. Notification must be given of all studies that use personal data (in whatever form). The Dutch Data Protection Authority (previously the CBP, College Bescherming Persoonsgegevens) website has more information on statutory obligations.
  - **Quality Assurance for Research Involving Human Subjects 2.0** (NFU, 2012)
  - **Medical Research (Human Subjects) Act** (WMO)
  - **Medical Treatment Contracts Act** (WGBO)

#### 2.2

If you perform research involving human subjects you must request **permission** to gather and/or use personal data and ensure the data are protected. Permission may take the form of **informed consent** or an **opt-out**. See for example the **CCMO consent form**.

To ensure the **reusability** of your data set, it is important that you request the consent of and inform your subjects of any **follow-up study** performed in addition to the original purpose for which data were collected.

Under the **Code of Conduct for Medical Research** no consent need be sought in a number of exceptional circumstances (see articles 5 and 6, 2004).

Under the **Personal Data Protection Act (WBP)** use of research data for **secondary analysis** is permitted on condition that it is not incompatible with the purpose for which the data were collected. See also the **Declaration of Helsinki on ethical principles**.
2.3 Two pieces of legislation apply to this issue:

- **WMO (Medical Research (Human Subjects) Act):** the researcher is responsible for ensuring the best possible protection of the subject’s privacy. In principle, research data must be used and stored anonymously.

- **The WBP (Personal Data Protection Act) applies to the processing of personal data** (in other words, actions involving personal data such as collection, recording, sorting, storage, updating, amending, etc.).

Under this legislation, and the **Code of Conduct for Medical Research based upon it**, anonymised or pseudonymised data must be used for research analysis.

- When data are **anonymised**, any details that can be linked to the individual or the healthcare institution must be altered in such a way that only anonymous details remain. These no longer qualify as personal data and are no longer subject to the WBP.

  **Anonymous data cannot be linked to an individual and cannot therefore be linked to data from another source.**

- When data are **pseudonymised all identifying data are encoded, so that identification of individuals can reasonably be assumed to be prevented or rendered impossible.**

  **When pseudonymised data are used, the possibility of linking data remains open. If you opt to pseudonymise data, you will often need the permission of the subjects to reuse the data. The data may be linked only once subjects have given their explicit consent. If the data are used for another purpose in the future (further use, or ‘secondary processing’), the researcher must again seek the permission of the subjects before linking the data.**

- A **trusted third party** can provide support with the exchange and accessibility of data sets containing privacy-sensitive information. See for example ZorgTTP.

**FAQs on the codes of conduct** (Federa) has more information on anonymous data and personal data.

3. Making data findable

3.1 Data sets can be found in **data repositories, catalogues (including metadata catalogues) and web portals**. They provide a list of the data sets in combination with descriptive information (data on data, or metadata).

A data repository is intended for the **long-term storage** of your data once your project is complete. For more information and answers to questions on this see part 6, *Reusable data and permanent storage*.

You are advised to choose a data repository at an early stage of your project. This will allow you to **consider the requirements set by the repository** when setting up your data collection, including a suitable file format and any metadata required.

If you are used to storing data in **international repositories**, you may continue to do so, particularly if this is a requirement stipulated by co-financiers or scientific journals. More information can for example be found on EMBL EBI. In this case, you need not also store your data in a Dutch repository.

(See also question 1.3)

3.2 Use a metadata standard for describing:

- the data set as a whole (this is what question 3.2 refers to);
- the individual records in the data set (this is what question 5.2 refers to).
This question (3.2) concerns metadata that provide information on: who gathered the data, under what circumstances, and what was measured/recorded.

Various metadata standards exist:
- generic or specialist
- detailed or general.
Ideally, the data description should give sufficient detail for potential users to decide whether the data are appropriate for their research.

To ensure data sets are findable and reusable, both now and in the future, metadata also have to be machine-readable, in accordance with the FAIR-data principles.

A repository may prescribe a certain metadata standard. If not, you are free to choose your own.

There are several things to consider when choosing a metadata standard:

- Choose a metadata standard specific to (appropriate for) your field. This allows you to produce a detailed description and enhanced the findability and reusability of data within the field. It also improves the potential for linking with other data sets.
- Examples of international standards can be found on DataCite. DataCite offers metadata standards for various types of data set. Every standard gives a number of data items that are needed to make the data findable. The data items can be completed with keywords or as free text.
- The Data Documentation Initiative Alliance is for the documentation of, among other things, social science and health research.
- Also choose a generic metadata standards so that the data collection can be ‘harvested’ by a data portal. The KNAW’s NARCIS portal is one example. The Dublin Core standards are another well-known example.

### 3.3
Use a persistent identifier (PI) to provide a permanent link to the data set. A PI is generated by a repository, and must be passed on to ZonMw at the end of the project.

Examples of PIs include DOI, Handle, URN and ARK.

The advantage of a persistent identifier over a normal link (URL) is that your data set will remain findable even if its location changes.

A data set can also be cited using a PI (data citation).

**More information about persistent identifiers and DOIs can be obtained from the International DOI Foundation (IDF), DataCite, TUDelft or the RDNL course.**

**Finally**
Do not store the persistent identifier with the identifying code *(see part 5, Interoperability)*. This is a code for the records in the data set.

### 4. Making data accessible

#### 4.1
It could be that an embargo period needs to be observed, during which your data will not be fully public, in connection with publication, public safety, privacy, intellectual property belonging to certain companies or commercial interests associated with exploitation of the research results, for example. ZonMw’s grant terms and conditions stipulate a 3-month embargo, and a maximum of 9 months for patents.

Explain the reasons for the embargo and how long it will last. If necessary, ZonMw will help you determine the duration of the embargo.
For more information see:
• Information on cofinancing on the ZonMw website.
• Guidance on property rights

4.3 In the event of restricted access you must give a precise description of the conditions that a research group seeking access to your research data must comply with.

In accordance with the FAIR guiding principles the legal status of the licences and the terms and conditions of use must be clear to ensure that data can actually be reused.

You can use internationally available standards, or draw up your own terms and conditions with the assistance of a lawyer.

The terms and conditions of use must be made available by your institute or department and may not be issued on an individual basis.

ZonMw requests that you ensure the terms and conditions cover as many of the criteria listed in the following question as possible.

4.4 See guidance on question 4.3

5. Making data interoperable

5.1 To make your research data interoperable it is advisable to use an IT standard (or data format) commonly used in your field for recording and coding your data. In other words, when choosing the standard, ensure that data are readable both by other researchers (unity of language) and by their computers (‘machine readable’ or ‘actionable’).

Your choice will also depend on the technical capabilities of your institute. You must therefore always seek the advice of your IT and/or data manager.

You can also obtain information from experts and/or their organisation’s website, such as:
• Nictiz: see ‘fourth layer of interoperability (application)’
• Examples of international standards can be found on DataCite. See also the guidance to question 3.2.
• Research Data Netherlands RDNL (or the participating institutes DANS, 3TU.Datacentrum, SURF),
• Dutch Techcentre for Life Sciences DTL
• If you use the Castor or Open Clinica data management system, your data are readable by and exchangeable with other computers.

These organisations provide information, IT services and facilities, IT standards and data storage, and they hold regular consultations with the IT departments of all Dutch research institutions.

ZonMw has made arrangements with these organisations concerning advice on data management in ZonMw projects.

5.2 To make your data interoperable, you are advised to use a metadata standard commonly used in your field (see Nictiz ‘third layer of interoperability’). Consider standard terminology, classifications and information standards.

Make good arrangements for this aspect of your study from the start of the project. It is much more work to do it in retrospect!

Commonly used metadata standards including:
• SNOMED CT
- **DataCite.** DataCite provides metadata standards for various types of data set.  
- The [Biosharing](https://biosharing.org) website lists metadata standards used in the medical sciences.  
- The [Castor EDC](https://www.castoredc.com) data management system uses recommended metadata standards.

**NB**

Question 3.2 also asks for a metadata standard, which refers to the description of the entire data set. This question, by contrast, concerns the exchangeability of data at the level of individual records. In order to link two similar data sets, the variables must be documented and coded in the same way.

### 5.3

Here ZonMw wishes to establish whether you have arranged privacy protection in such a way that your data cannot in principle be linked with other data sets.

In **research involving human subjects** certain ethical and legal requirements pertain to the linking of data. If the person has given their express **permission** for the secondary use of their data (including linking it with other data sets), the **identifying code** may be their BSN (Dutch tax and social security number), date of birth or postcode. If no permission has been given, the code must be a **pseudonym** that cannot be traced to the individual.

If the data are **anonymised** no permission is required, but it is not possible to link the data.

A trusted third party such as ZorgTTP can provide support with the exchange and accessibility of data sets containing privacy-sensitive information.

### 6. Making data reusable

#### 6.1

With a view to the reuse of data and possible replication of the research, you must archive at least the following **data and documentation**:

- the raw data (if you have reused existing data, you may have processed data rather than ‘raw’ data; in that case, the existing data must already be permanently archived, or must now be archived);
- the data on which publications are based;
- documentation on the research methodology used (such as code books, data manuals, metadata compilation, machine settings, use of SOPs, version management etc.), project proposal, approval from ethics committees such as METC, all stakeholders (researchers, laboratory assistants, test subjects etc.) in short, everything needed to ‘retrace the steps’;
- When storing **qualitative data**, describe the procedure used to transcribe the data (including conventions and symbols). Also look at [Atlas Ti](https://www.atlasti.com).

Together, this is known as a  **replication package**.

Details on this minimum set of information, focused specifically on research covered by the Medical Research (Human Subjects) Act (WMO), can be found in the publication "Quality Assurance for Research Involving Human Subjects 2.0" (NFU, 2012). Appendix 3 to this publication lists the documents that must be archived for the preparation, implementation and completion of this type of research.

#### 6.2

You must store research data for reuse or to verify and validate research results. This does not however apply to all data. In some situations it is easier or less costly to generate exactly the same data again than to store that same data. In some circumstances, it may also be in the interests of privacy to generate data again rather than storing it.

To help you determine which research data constitute valuable source material for further research, DANS has drawn up a checklist setting out ‘**General guidelines for**
the selection of research data for storage’. The guidelines set out the main reasons for storing research data in the long term. The checklist can be used by individual researchers, research groups, institutions, archive managers and funding bodies.

The Research Data Netherlands course (RDNL) also includes a user-friendly decision diagram for data selection.

6.4 In part 3 (Making data findable) you will have specified a repository where your data are findable. Is the repository of your choice suitable for permanent storage / long-term archiving, ensuring your data are permanently stored and reusable?

International guidelines are available for permanent long-term archiving. The simplest set of criteria is that of the international Data Seal of Approval (DSA). These criteria and the associated quality label are independent of subject area. A repository that stores data permanently is known as a trusted digital repository. In the Netherlands, CentERdata, DANS, 4TU.Datacentrum and SURF have this status.

There are barely any specific healthcare data repositories with the DSA. Researchers can however use an existing general repository (see the list of Seals on the website). You may also store and archive your data at your own institute, but please ensure maximum compliance with the guidelines.

If you choose a repository that does not have the Data Seal of Approval, DIN-31644-, ISO-16363- or WDS/ICSU certification, please make clear how you will ensure that the data remain effectively accessible and reusable after the project.

You are advised to ask your institute what possibilities it offers for permanent storage and archiving. To what extent does the storage facility operate in accordance with the criteria? And does it intend to apply for permanent storage certification? You might like to make the internal storage administrator aware of this option.

6.5 Different codes and guidelines specify different storage periods. ZonMw requests that you at least comply with the Dutch Code of Conduct for Academic Practice (VSNU): ‘Raw data must be stored for a period of at least 10 years’ (see also part 2, Legislation).

The publication ‘Quality Assurance for Research Involving Human Subjects 2.0’ (NFU, 2012) also specifies minimum storage periods for a range of material. (see question 6.1)

6.7 When estimating the costs you must distinguish between:

- (1) the costs of data management during the project and the preparation of data for archiving, and
- (2) the costs of permanent storage and archiving after the project.

Re. (1) These costs can be included in the budget submitted as part of your grant application. The costs of data management refer in part to:

- the time spent on metadata compilation and documentation during the study;
- the type of data and the size of the collection (you will have made an estimate of this in part 1. General features of the project);
- storage and securing of data during the project.

Re. (2) The costs of long-term storage will be borne by the institutes. A good picture exists of the costs of this and how they break down. They are at any rate determined by the type of data and the size (number of subjects and giga-/terabytes of the selected dataset (see selection process in questions 2 and 3)).

Several parties are working on a costing model for permanent data storage, though the knowledge to devise a suitable funding model for the storage of data is currently lacking. You can obtain information on the cost of archiving from the RDNL course.
4 Key items

When a project has ended, ZonMw checks whether the researcher has properly carried out the data management plan. This is done on the basis of a few key items, which are made public.

There are 5 key items:
1. The link to the repository's search engine, or to the online (metadata) catalogue or web portal listing the data collection
2. The DOI code (persistent identifier) providing a permanent link to the data collection
3. The link (or a persistent identifier) to the data collection's terms of use (this key item can be dropped if the collection is open access)
4. The metadata standard allowing the data collection to be linked with other collections
5. The link to the archive or trusted digital repository for long-term archival

Below, you'll see how these key items (marked in green) are listed in the questions of our DMP format.

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 My project's data collection will be findable for subsequent research</td>
<td>1. Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)  &lt;br&gt;2. Yes, it can be found through an online (metadata) catalogue or web portal (please specify)  &lt;br&gt;3. No, I have not yet chosen an archive or catalogue/web portal</td>
</tr>
<tr>
<td>(note: this is a key item, which you should report to ZonMw at the conclusion of your project).</td>
<td></td>
</tr>
<tr>
<td>3.3 I will be using a persistent identifier as a permanent link to my data collection</td>
<td>1. Yes, I will be using the DOI code  &lt;br&gt;2. Yes, in addition to the DOI code I will be using another persistent identifier (please specify)  &lt;br&gt;3. No, I will not be using a persistent identifier (please explain)</td>
</tr>
<tr>
<td>(note: this is a key item, which you should report to ZonMw at the conclusion of your project).</td>
<td></td>
</tr>
<tr>
<td>4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier)</td>
<td>1. Yes, my institution employs internationally available terms of use  &lt;br&gt;2. Yes, my institution has drafted a set of terms of use with the help of a legal advisor  &lt;br&gt;3. Not yet, my institution will draft a set of terms of use with the help of a legal advisor</td>
</tr>
<tr>
<td>(note: this is a key item, which you should report to ZonMw at the conclusion of your project).</td>
<td></td>
</tr>
<tr>
<td>5.2 I will select a metadata standard to allow my data collection to be linked to other collections</td>
<td>1. Yes, I will select a metadata standard from the list published by Biosharing (please specify)  &lt;br&gt;2. No (please explain)</td>
</tr>
</tbody>
</table>


6.4 I will select an **archive or repository** for (certified) long-term archival of my data collection once the project has ended

(note: this is a **key item**, which you should report to ZonMw at the conclusion of your project).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Yes, and this archive has a data seal of approval (please specify the archive)</td>
</tr>
<tr>
<td>2.</td>
<td>Yes, and this archive has a different form of certification (please specify the archive and certification)</td>
</tr>
<tr>
<td>3.</td>
<td>Yes, and this archive meets certification criteria and intends to get certified (please explain how your data will remain accessible and reusable in the long term)</td>
</tr>
<tr>
<td>4.</td>
<td>Not yet</td>
</tr>
</tbody>
</table>
5 Need help?

ZonMw supports researchers with data management within projects. If you have questions regarding the preparation of your data management plan, please contact the relevant program secretary or program assistant! They can assist you and answer your questions.