

Data Management Plan for Scientific Research

Data Management Plan for Scientific Research

Data Archiving and Networked Services (DANS)

Postbus 93067 | 2509 AB Den Haag

Anna van Saksenlaan 51 | 2593 HW Den Haag

+31 70 349 44 50

info@dans.knaw.nl | dans.knaw.nl



Introduction

Research data are valuable. They play an increasingly important role in science and should be effectively managed to ensure reuse and verification of research results. DANS stimulates sustainable access to research data and for this purpose provides [contracts](#)¹ in which agreements relating to managing research data are recorded.

Increasingly, a research project must set up a data management plan describing how research data are managed, documented and shared. This brochure contains focal points for drafting a data management plan and should be used especially in the initial stage of a data collection project. The final chapter contains the core elements of a data management plan.

This brochure has been drafted by using, amongst other things, the data management plans of the [UK Data Archive](#)² and the [Digital Curation Centre](#)³ in England, as well as of the [Interuniversity Consortium for Political and Social Research](#)⁴ (ICPSR) in the United States and the European [Horizon 2020](#)⁵ research programme. Experiences from the [Controlled Access to Research Data - Securely Stored](#)⁶ (CARDS) project were processed in the first version from April 2012. The November 2013 version benefited from experiences in the '[Regie in de Cloud](#)⁷' project. Various Dutch research institutes contributed to these projects. The minimum set of elements that should be included in a data management plan has been updated in this version – thanks to many examples.

The Hague, January 2015

1 See http://dans.knaw.nl/en/deposit/information-about-depositing-data/data-contract/data-contract?set_language=en

2 See <http://www.data-archive.ac.uk>

3 See http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

4 See <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html>

5 See http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

6 See <https://www.surf.nl/kennis-en-innovatie/kennisbank/2011/eindrapportage-project-cards-controlled-access-to-research-data-stored-securely.html> (in Dutch)

7 See <https://www.surf.nl/kennis-en-innovatie/kennisbank/2013/conclusies-en-resultaten-project-regie-in-de-cloud.html> (in Dutch)

Content

1. Setting up a data management plan

- 1.1 Why is a data management plan important?
- 1.2 When to write a data management plan
- 1.3 Stakeholders

2. Preparing the data collection project

- 2.1 General information about data collection
- 2.2 Overview of previously collected data
- 2.3 Options for available software and hardware
- 2.4 Establishing intellectual property and legal requirements
- 2.5 User information
- 2.6 Interoperability

3. Implementation of the data collection project

- 3.1 Data management
- 3.2 Options for available metadata

4. Archiving and making data available

- 4.1 Depositing data in a data archive or repository

5. The structure of a data management plan

- 5.1 Background and use of the template
- 5.2 Template for a data management plan

1 Setting up a data management plan

1.1 Why is a data management plan important?

There are various reasons for setting up a data management plan, including:

- Meeting grant requirements from the Netherlands Organisation for Scientific Research (NWO) or Horizon 2020, for instance
- Stimulating reuse by making the data easily findable, and explaining how others can gain access to the data
- Linking data to publications and other research information
- Enabling verification and replication of the research
- Very literally: planning and estimating all aspects of effective data management. This clarifies in advance the required data expertise and facilities, such as secure storage.

1.2 When to write a data management plan

The recommendation is to already consider data management when drafting a project proposal. Setting up the data management plan should at the latest take place in the initial phase of a data collection project. Starting with data management when a lot of research data have already been collected and stored is more difficult and costly, as it requires ample additional work. The plan may also still require alterations during the project, as research does not always go entirely according to plan. It is important to effectively document changes to the data management plan, so the researcher can account for these to stakeholders including funders.

1.3 Stakeholders

Who is involved with data management and its corresponding plan? First of all, obviously the researchers who will be conducting the research. Other stakeholders who play a role are management and support staff in the researcher's institution.

Why this is important

Involvement from the management and other key figures within the organisation is necessary for two reasons. Firstly, it is recommended to check the data management aspects where central policy, guidelines and support are desired and possible. Researchers ideally have room for filling in a data management plan, without having to work out all aspects ad hoc. Secondly, management should allow the researchers time to work on a data management plan. This is at present not a standard activity and experience demonstrates that the lead time can take several months. Once again, this suggests that data management should be organised at the level of an organisation, or part of an organisation. If an organisation has a data manager, this person should be involved in drafting the data management plan. The data manager can even take the lead, as he/she probably has, or will have, most experience with data management plans. This also applies to support staff in the institute's library. It is furthermore important to involve the IT department early in the process. This involvement may or may not be essential; the sooner this is clear, the better. Furthermore, the (external) data repository or data archive also prefers to know at an early stage about the nature, scale and access rights of future research data.

2 Preparing the data collection project

2.1 General information about data collection

It is important both for the data collection project's funder and the eventual project executor to formulate a general description of the data to be collected, their nature and scale; how the data will be collected or generated, and how the data will be made available in due course. Some disciplines may perhaps not call it 'data', but they do work with research data and results that are, along with publications, eligible for storage and availability. A data management plan is therefore also relevant in these disciplines.

2.2 Overview of previously collected data

If previously collected data exist that are comparable with the new data to be collected, it is important to create an overview of the data concerned and to explain why it is important to collect new data. This type of overview can be created by means of the catalogues of the main data archives and repositories.

Why this is important

The new data collection can be adapted to previously collected data. Taking stock of existing data can also save costs. Certain data may have been collected before and the required analyses can perhaps also be carried out on existing data.

Additional information

This type of overview can be created by means of the catalogues of the main data archives and repositories. In the Netherlands these organisations are, for instance, [Data Archiving and Networked Services](https://easy.dans.knaw.nl/ui/home)⁸ (DANS), [3TU.Datacentrum](http://datacentrum.3tu.nl/)⁹, the repositories of universities and the [Zorggegevens.nl](http://www.zorggegevens.nl)¹⁰ platform. The portal for research information, [NARCIS](http://www.narcis.nl/?Language=en)¹¹, also contains an increasing number of datasets. [re3data](http://re3data.org/)¹² provides an international registry of data archives and repositories.

2.3 Choice of software and hardware to be used

It should subsequently be determined which hardware and software will be used. When choosing software, it is important to consider standard discipline-specific data formats.

Why this is important

Obsolescence of hardware and software means that data reuse is eventually no longer possible, endangering the digital sustainability of data. It is therefore recommended to use software that in the long term can be converted into a still usable file format by a data archive or repository.

8 See <https://easy.dans.knaw.nl/ui/home>

9 See <http://datacentrum.3tu.nl/>

10 See <http://www.zorggegevens.nl> (in Dutch)

11 See <http://www.narcis.nl/?Language=en>

12 See <http://re3data.org/>

Additional information

DANS has carried out research on the sustainability of file formats and drawn up a [list of preferred and accepted formats](#)¹³ of the following file types:

- Text files, such as PDF/A, plain text and mark-up, such as HTML
- Spread sheets
- Statistical files
- Databases
- Visual material
- Audiovisual files
- Cartographic data
- Geographic Information System (GIS data)

2.4 Establishing intellectual property and legal requirements

For establishing intellectual property and legal requirements, it is important to find answers to the following questions: which person or institution owns the copyright of the collected data, and will these rights in due time perhaps be transferred? Are there legal requirements or conditions for sharing and archiving the data? Will copyright-protected material be used, for instance pictures, certain measuring instruments or scales? If so, how does the project obtain permission to use and spread the material?

Why this is important

To be able to make the data available via a data archive or a repository upon completing a data collection project, it is essential to know who owns the intellectual property rights. If copyright-protected material is used in collecting data, this copyright should be respected.

Additional information

Agreements apply to depositing and using data, based on Dutch and European legislation and on the '[Gedragscode van de Vereniging van Universiteiten](#)¹⁴ (Code of Conduct of the University Association, VSNU) for the use of personal data in scientific research. This code is an elaboration of the Dutch Personal Data Protection Act.

For depositing data from social and behavioural sciences, DANS identified various legal aspects regarding data depositing and making data available. The regulations are described in more detail in the Licence Agreement and the Conditions of Use. The Licence Agreement contains the agreements between the data depositor and DANS. The Conditions of Use describe the agreements between DANS and the data's users. Although the information specifically applies to social and behavioural sciences, it also includes elements that apply to other scientific disciplines. You can find more about DANS legal information on the [website](#)¹⁵.

13 See http://dans.knaw.nl/en/deposit/information-about-depositing-data?set_language=en

14 See <http://www.vsnul.nl/code-pers-gegevens.html> (in Dutch)

15 See http://dans.knaw.nl/en/about/organisation-and-policy/legal-information?set_language=en

2.5 User information

Are measures taken to gain insight into the use of the data? For instance, who will use the data and for what purpose?

Why this is important

The users for whom the data are collected can be decisive for the way in which data are managed, documented and shared.

2.6 Interoperability

It is important in many disciplines that the data can be linked to data managed by an external organisation, and also to data collected or generated in the future. Which standards or agreements concerning data description exist within the discipline or the knowledge institute? What does the research funder prefer?

Why this is important

This information is essential for being able to make specific agreements – at an early stage – with the intended data managing institutions.

3 Implementation of the data collection project

3.1 Data management

When executing the data collection project, it is important to establish who is mainly responsible for data management, and who is responsible for documenting all phases of the data collection project. Is a standard naming convention concerned? How, if applicable, is confidential handling of personal, commercial and military data ensured? How and where are copies of the data stored, and how is secure storage of this data guaranteed? How is version management organised? How is access to data organised for project members within, and outside of, their own institution during each phase of the data collection?

Why this is important

Answering the above questions ensures that data management is adequately organised during the research phase. Furthermore, digital data are vulnerable and the best method is to store multiple copies in multiple locations. Securely storing sensitive data, if applicable, also plays an important part. The researcher can usually rely on his or her organisation for information about these provision types.

Additional information

Several prominent data archives, such as the Interuniversity Consortium for Political and Social Research (ICPSR) in the United States and the UK Data Archive in England, provide extensive advice and templates for data management plans. In addition to its template, the British Digital Curation Centre furthermore provides references to filled-in plans. The working group of Dutch university libraries and the Royal Library (UKB working group) has also inventoried many templates. For detailed information on various aspects of proper data handling, please check the following links:

[UK Data Archive – Managing and sharing data – Best practice for researchers](#)¹⁶
[ICPSR – Elements of a Data Management Plan](#)¹⁷
[Australian National Data Service – Data Management for Researchers](#)¹⁸
[Digital Curation Centre – Data Management Plans](#)¹⁹
[UKB work group – Report Data Management Plans](#)²⁰

16 See <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf> or <http://www.uk.sagepub.com/books/9781446267264>

17 See <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/resources.html>

18 See <http://www.andis.org.au/researchers/manage-data.html>

19 See <http://www.dcc.ac.uk/resources/data-management-plans>

20 See <https://wiki.surfnet.nl/display/RD/Datamanagementplannen>

3.2 Choice of metadata to be used

Which metadata will be used, and do they belong to a particular metadata standard?

Why this is important

Metadata – data on the research data themselves – are necessary for finding data in an (online) catalogue. Adding rich, detailed metadata furthermore implies that the content of a data collection is adequately accessible. Thanks to metadata those who are interested can better determine if a data collection is relevant and usable for their own research. Metadata are thus essential for reuse.

Additional information

Different scientific disciplines have their own metadata standards; these can be intricate like the Data Documentation Initiative specification for social and behavioural sciences, or record the minimum required information, such as the Minimal Information (MI) Standards in life sciences. There are also generic metadata standards, such as Dublin Core. Some types of metadata provide editors for assigning metadata to data. The Digital Curation Centre (DCC) refers to a large number of generic and discipline-specific metadata formats.

[Dublin Core Metadata Standard](#)²¹

[Data Documentation Initiative](#)²²

[Metadata Geo Data](#)²³

[Minimum Information Standards](#)²⁴

[NESSTAR](#)²⁵

[DCC – Metadata Standards](#)²⁶

21 See <http://dublincore.org/specifications>

22 See <http://www.ddialliance.org/>

23 See <http://www.geonovum.nl/onderwerpen/metadadata?tab=standaarden> (in Dutch)

24 See http://en.wikipedia.org/wiki/Minimum_Information_Standards

25 See <http://www.nesstar.com/>

26 See <http://www.dcc.ac.uk/resources/metadadata-standards>

4 Archiving and making data available

4.1 Depositing data in a data archive or repository

How and where will data be stored for the long term? Indicate how data will be selected for archiving, and how long data should be kept for. Are the costs of archiving at a data archive or a repository included in the budget of the data collection project?

Why this is important

It is important to be able to make data available for new research; access to research data also improves the openness and verifiability of scientific research. Data archives and repositories impose certain requirements on data and its documentation for reuse purposes. This is also important for reuse in the long term. It is therefore recommended to reserve part of the research budget for this.

Additional information

It is recommended to deposit data at a data archive that meets the guidelines of the internationally established Data Seal of Approval. The Data Seal of Approval (DSA) defines requirements for the trustworthiness of data archives. The guidelines are a first step in a series of international certification initiatives. The DANS and 3TU.Datacentrum repositories, which have joined forces in [Research Data Netherlands](#)²⁷ (RDNL), both carry the Data Seal of Approval. Instructions for online depositing in the DANS EASY archive can be found on the [DANS website](#)²⁸. This page furthermore contains information on guidelines for depositing archaeological data. Instructions for online depositing in the 3TU.Datacentrum data archive can be found on the [3TU.Datacentrum website](#)²⁹. Further information about the Data Seal of Approval, including the overview of allocated Seals, can be found on [the DSA website](#)³⁰. Additional information about the obligation to deposit data from projects granted by, for example, the Netherlands Organisation for Scientific Research (NWO) or the government can also be found on the [DANS website](#)³¹.

Publishers of scientific journals are increasingly focusing on storing research data. This frequently concerns the data on which a publication is based, as well as so-called supplementary data. The Data Availability Policy (DAP) of a journal or publisher can contain conditions for authors with regard to making data available, or can refer to data storage possibilities. The British JoRD project ([Journal Research Data Policy Bank](#)³²) offers examples of data availability policies in social sciences, whilst the Dutch [CLIO-DAP project](#)³³ focuses on the influence field of researcher, editorial staff, reviewer and repository.

27 See <http://www.researchdata.nl/en/>

28 See http://dans.knaw.nl/en/deposit/information-about-depositing-data?set_language=en

29 See <http://datacentrum.3tu.nl/en/what-we-offer/upload-your-data/>

30 See <http://www.datasealofapproval.org/>

31 See http://dans.knaw.nl/en/deposit/information-about-depositing-data/data-contract/data-contract?set_language=en

32 See <http://jordproject.wordpress.com/project-data/social-science-journals-that-have-a-researchdata-policy/>

33 See <http://xposre.nl/cliodap/>

The recommendation of depositing data at a certified archive or repository also remains in force in these developments.

In 2014, the VSNU in the [Nederlandse Gedragscode Wetenschapsbeoefening](#)³⁴ (the Netherlands Code of Conduct for Academic Practice) set the minimum retention period of raw data at ten years.

34 See

[http://www.vsnunl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20\(version%202014\).pdf](http://www.vsnunl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20(version%202014).pdf)

5 The structure of a data management plan

5.1 Background and use of the template

Finally, this brochure contains a template with the main elements of a data management plan. The template is discipline-independent and intends to strengthen data awareness of researchers, as well as lay the foundation for a usable plan that can be adapted during research. The selection of these core elements is mainly based on the following sources:

- The checklist for a [data management plan from the Digital Curation Centre](#)³⁵ (version 4, 2013)
- The questions for a data management plan from [Horizon 2020](#)³⁶
- The NWO memorandum 'Project team Data management' (2013)
- The [advice 'Data: Digitale diamanten'](#) (Data: Digital Diamonds, 2013) from RIVM, NIVEL, DANS, CentERdata and BBMRI-NL for ZonMw³⁷
- The ['Research Data Management' report](#) from the UKB working group³⁸, which has inventoried over fifty templates from the Netherlands and abroad (2014)
- Feedback on previous versions of this DANS brochure.

Suggestions for those who want to process these core aspects into their own template:

- For a data section in a grant application or project proposal it is recommended to make a selection from the list below. This phase does not require writing a complete data management plan, nor should a funder ask for a complete plan. However, a data section is a valuable instrument for drawing attention to data management.
- Various elements from the list can be further clarified. This preferably takes place using examples from the discipline or institution of the researcher involved.
- Templates and instructions come in many forms: as a list of questions or focal points, as a template that can be filled in by the researcher, as a checklist with statements for which the researcher can select *Yes/No/Not applicable*, or as a combination of these structures. There are not yet indications that one of these forms is more effective.

35 See http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

36 See

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

37 See

http://www.zonmw.nl/fileadmin/documenten/Toegang_Tot_Data_TTD/TTD_Eindrapport_Datamanagement.pdf (in Dutch)

38 See <https://wiki.surfnet.nl/display/RD/Datamanagementplannen> (in Dutch)

5.2 Template for a data management plan

1	Administrative information	
1.a	Project title, principal researcher, funder(s), date of this plan and of previous versions	
1.b	Who is primarily responsible for data management?	
2	Describing the data	
2.a	Are existing data reused or are new data generated?	
2.b	What type(s) of data is concerned; file size; growth rate?	
3	Standards and metadata, or everything required to find and use the data	
3.a	Which metadata standards are used (findability)?	
3.b	Which coding etc. is used for enabling future linking to other data (interpretation, interoperability)?	
3.c	Which software and hardware are used (interpretation, usability)?	
3.d	What is documented and stored to enable replication?	
4	Ethical and legal	
4.a	In collecting or generating data, how is required permission obtained from the data supplier/ test subjects / ...? Which restrictions, if any, apply during research?	
4.b	How are sensitive data protected during and after the project?	
4.c	What are the agreements in the case of stakeholders leaving (early)?	
4.d	Are data available Open Access after the project, perhaps after an embargo period? If not, which conditions apply?	
5	Storage and archiving	
5.a	How is sufficient storage and back-up capacity organised during the project, including version management? Are the expenses for this covered, and if not, ...?	
5.b	Where, and how long after completing the project, are data available for follow-up research and verification? Is this a Trusted Digital Repository with an international certification? If not, how will the data be findable, sustainably accessible and usable? Consider metadata and persistent identifiers, such as DOI, Handle and URN.	
5.c	Are the expenses for (preparing the data for) archiving covered?	



Data Archiving and Networked Services (DANS)

DANS promotes sustained access to digital research data. For this, DANS encourages scientific researchers to archive and reuse data in a sustained form, for instance via the online archiving system EASY (easy.dans.knaw.nl) and DataverseNL (dataverse.nl). With NARCIS (narcis.nl), DANS also provides access to thousands of scientific datasets, publications and other research information in the Netherlands. The institute furthermore provides training and consultancy and carries out research on sustained access to digital information. Driven by data, DANS ensures the further improvement of access to digital research data with its services and participation in (inter)national projects and networks. Please visit dans.knaw.nl/en for more information and contact details.

Data Archiving and Networked Services (DANS)

Postbus 93067 | 2509 AB Den Haag
Anna van Saksenlaan 51 | 2593 HW Den Haag
+31 70 349 44 50
info@dans.knaw.nl | dans.knaw.nl

DANS is an institute of KNAW and NWO



Driven by data