



university of
 groningen

faculty of behavioural
 and social sciences

nieuwenhuis institute

Research Data Management Protocol for Nieuwenhuis Institute Researchers

7 November 2017

Version: 2.0

Contents

1	Publication packages: good scientific practice	5
2	Who, What, Where, and When?	7
2.1	Introduction: the default procedure and how to deal with it	7
2.2	Who?	7
2.3	What?	8
2.4	Where?	8
2.5	When?	9
3	Publication packages	10
3.1	General information	10
3.2	Content of publication packages	10
3.3	Folder structure for projects	12
3.4	PhD students	12
4	Examples of publication packages	13
4.1	Example of a qualitative publication package	13
4.2	Example of a quantitative publication package	16
	Appendix 1. Checklist for publication packages	18
	Appendix 2. Sustainable data formats	19
	Appendix 3. Read me first metadata file template	21

1 Publication packages: good scientific practice¹

Scientists publish studies because they think they have something interesting to report. In order to evaluate the strength of a study, including the design, the materials, the data and the analyses, the consumers of science should be able to check which choices were made. Typically, the information in published papers is insufficient to deduce these choices, and although few disagree that transparency about these steps is a scientific virtue, it is still far from standard practice.

Recent developments have given rise to more attention for transparency. An increasing number of grant suppliers, such as NWO, KNAW and ZonMw, and scientific journals demand that researcher think about data management and that research data are made accessible for other researchers in so-called trusted repositories. The call for transparency in science is by no means new. However, at least partly due to the discovery of a few fraudulent scientists, there was a sudden increase in the attention for topics like transparency and openness, despite the feeling that these cases were the exception rather than the rule. Some papers showed that more subtle ways of unwanted biases in science were far more common than many expected. A few recent large-scale replication projects² showed that the mere fact that a study passed peer-review was no guarantee that its results were replicable: in fact, only 36 out of 97 studies found a clear effect in the same direction.

Not only is it our moral obligation to be open about our findings, but it is also in all of our interest to rebuild trust in science by many in and outside academia. Many researchers initially might be sceptical about plans to regulate transparency, because it seems to increase the bureaucratic burdens, and it might feel like your integrity is implicitly questioned. Additionally, many fear that documenting data increases the chance of being scooped (that is, other researchers trying to out-publish you with your data).

Increased transparency and data management, however, is not only in the interest of the scientific community as a whole, but also of the researcher who documents his or her research choices. First, it increases the replicability, credibility, and hence the reputation of the researcher, provided that the documenting is done well. Moreover, it is a backup that protects us against loss of files by crashed laptops, stolen computers, or the moving of one of the authors to another university. Moreover, many grant suppliers and journals nowadays request a proper documentation, and if this is already standard, it doesn't have to be done again. It is to be expected that the push for more transparency will only increase in the next years and potentially cause papers without proper documentation unpublishable in the near future.

Thus, the call for more transparency seems well-founded, and in the interest of science, society and the individual researchers to maintain a FAIR data management (FAIR: Findable, Accessible, Interoperable, Reusable).³ In the sequel of this Research Data Management Protocol (RDMP), it will be clarified how researchers of the Nieuwenhuis Institute should deal with data management.

In summary, the most important reason to document research and store data in a safe place is that the author and co-authors and, when applicable, supervisors will still have access to the data and know which important steps were taken during the research in a few years (i.e., **safekeeping** of valuable information). In addition, if needed, others can **verify** the research steps taken and, if applied to the scientific method used, **replicate** the research.

¹ This protocol was formulated by prof. dr. J.J. H. Dekker, dr. R. Hoekstra, dr. D.D.N.M. Kostons and R. R. De Leeuw, MSc, who were supported by dr. S.W. de Boer. We thank dr. B. Oldenburg (Department of Sociology) for the permission to make use of the *Publication Packages & Research Packages Protocol* for sociology from February 2017.

² E.g. Brian A. Nosek, Alexander A. Aarts, Christopher J. Anderson, Joanna E. Anderson, and H. Barry Kappes (2015). Estimating the reproducibility of psychological science. *Science* 349 (6251), aa4716.

³ <https://www.nwo.nl/beleid/open+science/datamanagement>

This protocol for Nieuwenhuis Institute researchers builds on two years of experience in the Nieuwenhuis Institute by PhD-students. Following the guidelines of the University of Groningen⁴ and the Association of Universities in the Netherlands (VSNU)⁵, the Nieuwenhuis Institute then implemented policy to stimulate PhD-students to handle their research data in a responsible way.⁶

From January, 1, 2018, *all* Nieuwenhuis Institute researchers are requested to document important steps that are taken during their research and to store their digital data in special folders or *packages*. The Nieuwenhuis Institute works with one type of package: the publication package. **Publication packages** contain all raw data, syntaxes, and other important information needed to follow and track the methodological steps and choices made and when possible replicate the research that is published in a scientific journal, book (chapter) or research report. As further illustrated in chapter 3, publication packages that are part of a larger whole, such as a thesis or a project, will be organized in overarching folders.

In principle, the packages are intended to verify and provide sufficient information for replication purposes or meta-analyses. They are not meant to share (raw) data for reuse. However, the policy of the Nieuwenhuis Institute is that (raw) data should become publicly available if possible. This is the *default option*. Whenever there are legal, ethical, scientific or contractual reasons not to do so, it is up to the researcher to explain why this default option could not be met.

⁴ <http://www.rug.nl/research/search/research-data-office/policy/documents/2015-research-data-policy.pdf>

⁵ http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf

⁶ *Data Management Protocol of the Nieuwenhuis Institute*, 2016.

2 Who, What, Where, and When?

2.1 Introduction: the default procedure and how to deal with it

The default procedure is that the datasets used for the research (see further on datasets section 'What' of this chapter) will be put into the publication packages.

Researchers when preparing their publication packages always have to make certain if this default procedure can be used fully for their specific package. Therefore, they have to consider three main reasons mentioned below – concerning research methods, legal, and ethical reasons – and to explain and to justify if, and so why, a departure from the default procedure is necessary. The three reasons are:

- 1) The specific research method used prevents to follow *in toto* the default procedure;
- 2) Legal reasons prevent to follow *in toto* the default procedure;
- 3) Ethical reasons prevent to follow *in toto* the default procedure.

Ad 1) Reasons related to the method used

Research at the Nieuwenhuis Institute is diverse in its scientific methods which also can result in diversity in the collection and character of data. The research can be confirmatory or exploratory, interpretive or testing, quantitative or qualitative, producing new data or using existing data, like CBS or DANS data, data in archives, and in collections of films, video's, photographs, paintings, drawings, or other image material), and also using mix-methods. What has to be stored in a publication package is therefore also dependent on the methods used. Important is that the package makes it possible to verify (follow the main steps made by the researcher), and when possible to replicate the research.

Ad 2) Legal reasons

Several legal reasons might prevent uploading (parts of) raw data, such as issues concerning privacy (for example recognisability in videos), ownership (copyrighted materials such as materials stored elsewhere), or contractual obligations (for example safe-storing of information). All these issues can limit or even prevent putting datasets in DMPs. Such deviations from regular practice should be put in a memo in the publication package or parent folder.

Ad 3) Ethical reasons

Even if there are no legal considerations that prevent making (all) the data available, privacy issues also have an ethical aspect that might prevent storing the data in a publication package. Specific types of research that yield data that are privacy-sensitive should never be uploaded for a publication package, among them data on the BSS' own staff or students, as these are more prone for internal misuse. It might also be possible that specific participants do want to partake in a study, but do not want their raw data available stored on a server. Any deviations from regular practice should be put in a memo in the publication package or parent folder.

2.2 Who?

All members of the Nieuwenhuis Institute (**staff**, **postdocs**, and **PhD students**⁷) are responsible for ensuring that there is a publication package for every accepted article in a

⁷ The supervisors have the final responsibility for publication packages of PhD-projects, while its realization is the responsibility of the PhD-student.

scientific journal, book (chapter) or scientific report of which they are the first author. For more information about the publication packages, see chapters 3 and 4.

Access rights

By default, only the owner of the **package** has reading and writing rights for that package. In the case of multiple packages in a single folder, the owner of that folder (such as a supervisor or project leader) will have reading rights to all publication packages within that folder (and all subfolders). The coordinator – the Nieuwenhuis Institute Research Director and member of the management team of the PEDOK-department - and the Technical Support Facilities director also have reading rights.

It is important that others involved in the research, among them co-authors, research colleagues and supervisors, can access the packages as well. Access can be requested via this form:

<http://myuniversity.rug.nl/infonet/medewerkers/ict/serviceesk/mutation-y-drive>

Access rights need to be requested per package. At present it is not possible for co-authors/research colleagues/supervisors from other universities to access the packages. Developments in ICT and the option of an f-account should provide the opportunity to do so in the near future.

2.3 What?

A package should contain all raw (quantitative and qualitative) data⁸ used in a certain study. When it is not possible for certain reasons to include the data in a package (ethical guidelines, insurmountable logistical problems because of specific research methods used, privacy-related reasons or formal property rights), the responsible researcher is expected to add this information as a Word file to the publication package and contact the coordinator, i.e. the Nieuwenhuis Institute Research Director.

As a rule, the publication package should *never* contain data that are not anonymized or pseudonymized. In some cases it will be not possible to anonymize or pseudonymize the raw data directly. Then, the data will be labelled as personal data. Examples are video material, photos, audio tapes and informed consent forms. In these cases, the publication package should contain the first step in the analysis of that data which can be anonymized or pseudonymized, such as a transcript in the case of audio recordings. It is important to note that the **original data** still needs to be safely stored, not on the Y drive (see further section 2.4.) but elsewhere. For the **digital storage of those personal data** the technicians of the **BSS Technical Support Facilities** can help you.

Note that when the researcher makes changes to already uploaded data, this should be noted in the Word file as well. In chapter 3, more information will be given about the publication packages; chapter 4 contains some worked out examples.

2.4 Where?

Every member of the Nieuwenhuis Institute has access to the now safest digital environment of the university, namely the Y drive. On this Y drive, a folder will be created for publication packages (Y:\Staff\gmw\.....). Margreet Degen-Klabou (m.r.degen@rug.nl) can be requested for the creation of your folder. The publication package should be placed in the folder with the researcher's P-number, which can be found in the folder of the research cluster (e.g., De Leeuw, see example in chapter three). Contact Margreet Degen-Klabou (m.r.degen@rug.nl) also when

⁸ Note that with 'raw data' we mean (primary and secondary) empirical data, but also other types of data, such as meta-analyses.

the folder with your P-number cannot be found, for instance, when someone recently started working at the department. And contact the coordinator when you want to make a new package.

All data on the Y drive are retained for at least ten years.⁹ The Y drive backup system makes sure that everything stored on that drive can be retrieved, even if someone edits or deletes data after it has been saved). When someone leaves the department he or she can request to keep his or her P-number for access to the packages.

2.5 When?

It is wise to make the package as soon as possible after the completion of the research project. Note that some grant suppliers request a data management paragraph or protocol and the researcher is required to realize the setup of the publication package within a specific time after the grant approval.

In any case, the package should be ready **within three months** after the publication or completion of the research project.

⁹ This period follows the recommendation of the *Richtlijn archivering wetenschappelijk onderzoek voor Nederlandse faculteiten Maatschappij-en Gedragwetenschappen* from 2017.

3 Publication packages

3.1 General information

Publication packages are made for every published scientific article or book (chapter) of which a member of the Nieuwenhuis Institute is the **first author**. This author is responsible for ensuring that there is a publication package. Such a package contains all data, syntaxes, and other important information belonging to the published paper.

On the Y drive a folder has been created for the publication packages (Y:\Staff\gmw\...). The publication packages should be placed in the folder with your P-number which can be found in the folder of your research cluster (e.g., De Leeuw). The name of the publication package should be 'PUB' followed by the surnames of the first, second, and last authors and year of publication (e.g., ***PUBDeLeeuwDeBoerBijstraMinnaert2017***)

3.2 Content of publication packages

Given the different types of research within the Nieuwenhuis Institute (see Chapter 2, 'Introduction: the default procedure and how to deal with it') it is impossible to provide an exhaustive list of information that should be included. When creating a publication package, you could keep in mind that a colleague who was not involved in the project should be able to understand, follow and track the methodological steps and choices and, if possible given the methodology of the study, be able to reproduce the published research. Figure 3.1 and Table 3.1 provides an overview of the most used folders in a publication packages.

Figure 1. Example of a folder overview publication package

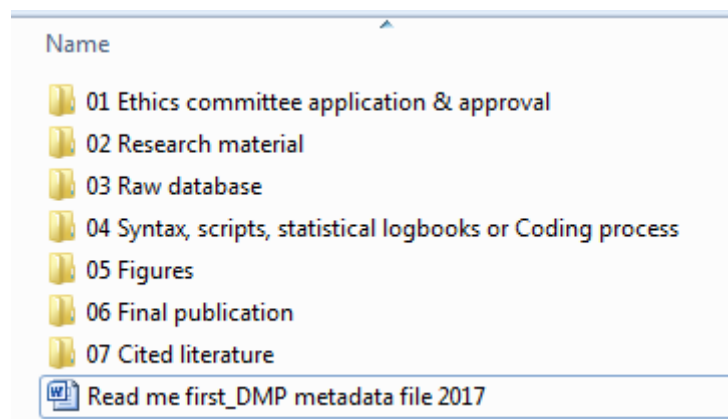


Table 1. Most used folders and description of content, publication package

Folder	What
Cited literature	<ul style="list-style-type: none"> • Document with references to cited books, chapters, data bases, websites etc. • References to all cited articles
Ethics committee application & approval	<ul style="list-style-type: none"> • A description of whether the project was evaluated by an ethics committee (if so, include application and letter of approval)
Figures	<ul style="list-style-type: none"> • Published figures
Raw database	<ul style="list-style-type: none"> • Description of sample/participants • Raw (or original) digital data¹⁰. Contact the coordinator when it is not possible to store the data in the publication package (e.g., due to ethical guidelines, insurmountable logistical problems in relationship to the research methods used, privacy-related reasons or formal property rights) • In the case of non-digital data, a document that states where the data are stored and how they can be accessed (by referring to inventory details)
Research material	<ul style="list-style-type: none"> • Overview of questions that were asked (i.e., a questionnaire) • Explanation of the variables (i.e., a codebook)
Syntax, scripts, statistical logbooks	<ul style="list-style-type: none"> • Syntax or statistical logbooks of the processing of the raw data and a properly documented and edited database to enable analyses such as those reported in the article to be replicated from the raw database • Syntax for created variables/scales or coding schemes for qualitative data • Other digital relevant research material, like stimuli, instructional texts, experiment leader protocols, video material, software for simulation studies, ...
Final publication	<ul style="list-style-type: none"> • The published articles or books (chapters)
Other	<ul style="list-style-type: none"> • A <i>readme</i> file (see example read me file publication packages)

It is important that others involved in the research (co-authors/supervisors) can access the package. Access can be requested via this form:

<http://myuniversity.rug.nl/infonet/medewerkers/ict/servicedesk/mutation-y-drive>

The package should be ready within three months after the article or book(chapter) is published.¹¹

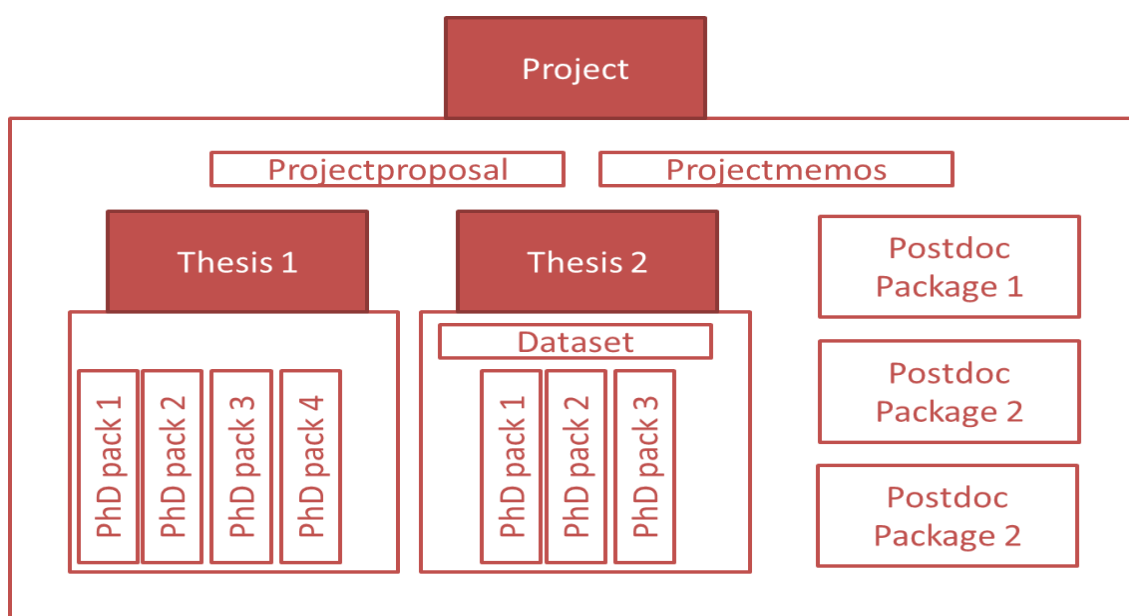
¹¹ Cf. *Richtlijn archivering wetenschappelijk onderzoek voor Nederlandse faculteiten Maatschappij-en Gedragwetenschappen* (2017).

3.3 Folder structure for projects

A publication package contains all data and other information relevant to a single publication. However, publications are often not singular entities but are part of a larger project. This gets structured in folders.

Take for example the fictitious project “For the Win”, in which a professor has successfully acquired a five year research project, with two PhD students and a Postdoc. This means two dissertations and several separate studies are present in the folder. One of the PhDs is focusing on gathering data for each study, whereas the other PhD is using a longitudinal dataset used across the various studies. This is why (see figure 2) the folder for thesis 1 includes all of its information and data in each separate pack, whereas the folder for thesis 2 has its dataset outside of the three publication packages (and the syntax for each of the three studies within the package).

Figure 2: Folder structure for projects



Folders are coloured in, packages and loose files have a blank background.

3.4 PhD students

The policy of the university is that PhD students can only defend their thesis when there is a publication package for each chapter of their dissertation (excluding the introduction and conclusion). This implies that the graduate school will not send out the thesis to the reading committee before the packages are in place.

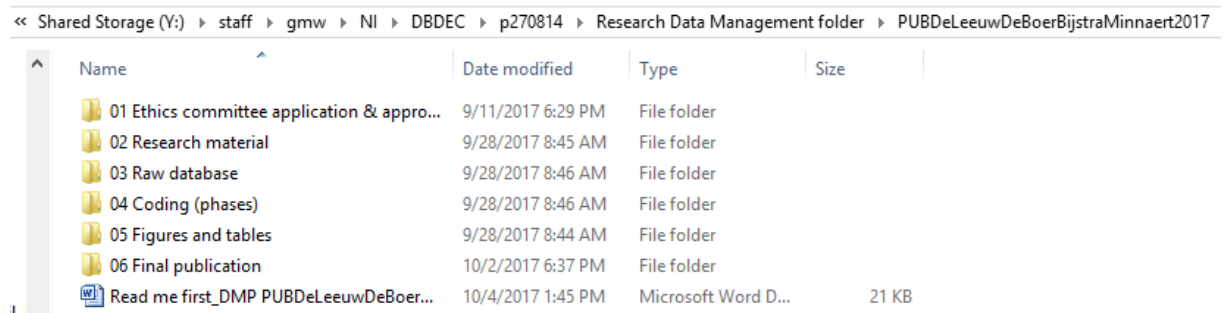
For PhD students the same guidelines as described in paragraph 3.1 apply. Please make a separate folder for each chapter of the dissertation (excluding the introduction and conclusion) and clearly indicate the chapter of the thesis in the name of the folder (e.g., DeLeeuw_chapter1). Include a Word or PDF version of the chapter, also when the chapter is not yet published. Send an e-mail to the coordinator when you have completed the packages. The coordinator will check whether all necessary files are there and inform the graduate school.

4 Examples of publication packages

In this chapter two examples of publication packages are provided, one of a qualitative study/publication and one of a quantitative study/publication. As a matter of fact, every study has different characteristics that will lead to differently structured publication packages, often resulting from a mix of qualitative and quantitative methods used. Therefore, the examples should by no means be seen as the standard: it is up to the researcher to tailor the package to the characteristics of his or her study.

4.1 Example of a qualitative publication package

This example of a qualitative publication package is from a sub-study of a single PhD project. The image below is a screenshot of the folders within the publication package and contains raw data, final data set, methodology descriptions and process, figures and tables published in the article, the Word version and Pdf version of the article.



Name	Date modified	Type	Size
01 Ethics committee application & appro...	9/11/2017 6:29 PM	File folder	
02 Research material	9/28/2017 8:45 AM	File folder	
03 Raw database	9/28/2017 8:46 AM	File folder	
04 Coding (phases)	9/28/2017 8:46 AM	File folder	
05 Figures and tables	9/28/2017 8:44 AM	File folder	
06 Final publication	10/2/2017 6:37 PM	File folder	
Read me first_DMP PUBDeLeeuwDeBoer...	10/4/2017 1:45 PM	Microsoft Word D...	21 KB

Read me firstDMP file

This document is the read me first metadata file describing what is in each folder. In Appendix 3 there is a format provided for a metadata file document. Using this file provides steps preventing that you can miss a folder, file or description of your stored data. For the example publication the format of Appendix 3 is used.

Read me first metadata file folder PUBDeLeeuwDeBoerBijstraMinnaert2017

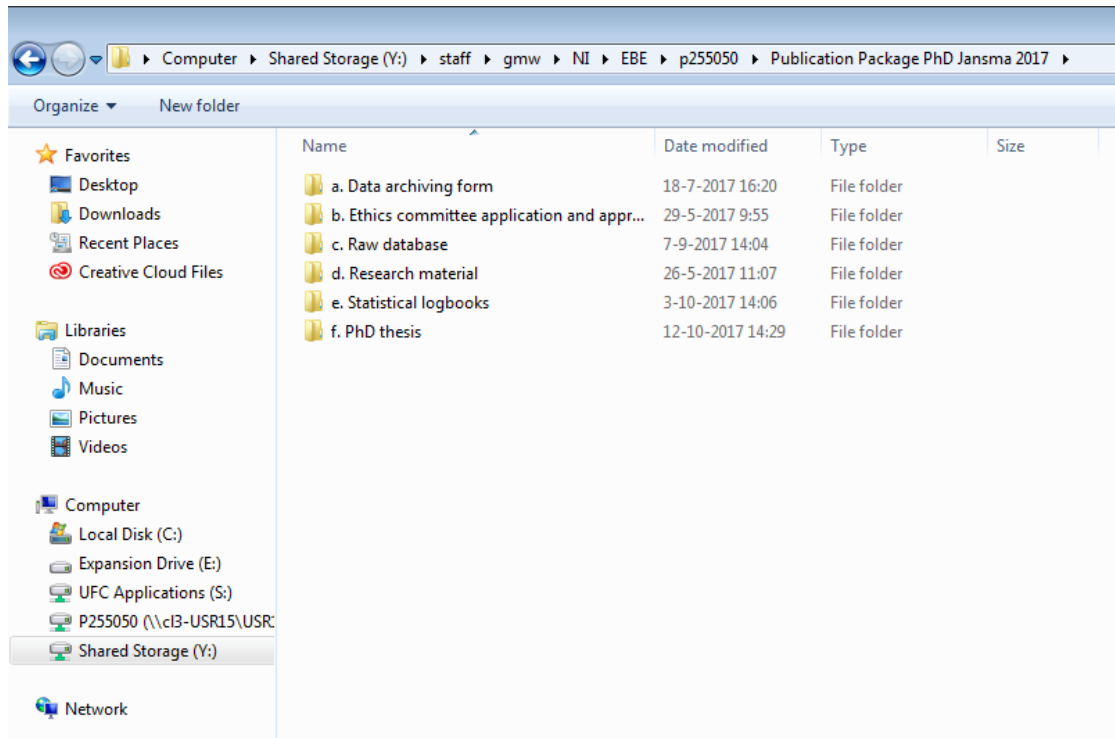
Administration	
Name	Renske de Leeuw
P-number	P270814
Name of research group	Special Educational Needs and Youth Care
Name of project	The needs for social participation of students with Social Emotional Behavior problems in the regular education
Description of the study	Qualitative study with focus group; expert panels with primary education teachers about stimulating the social participation of students with social and emotional difficulties and behavioral problems.
Funding agency	Stichting Kinderpostzegels Nederland under Grant 8508165
Institute project code	170460112
Project duration	PhD project April 2014 – November 2019
DMP version	Nieuwenhuis Institute Manual 2017

Project data managers <i>For every data manager: Name, contribution/description what is done regarding data collection</i>	n.a.
External collaboration	Yes, with Jan Bijstra from the Stichting Kinderpostzegels Nederland under Grant 8508165
Data collection	
Permits & licenses <i>Describe which type of informed consent was given, by whom, how and where consent forms are stored.</i>	<p>Approval for this research was granted by the ethical committee <i>Pedagogische Wetenschappen en Onderwijskunde</i>, at June 27 2013. (See document in folder 01 Ethics committee application & approval)</p> <p>For all 41 participants, active informed consent was obtained from all participating teachers in the focus groups.</p>
Description of the raw data <i>Describe what kind of raw data is collected</i>	<p>The focus groups followed the principles of the incident method of Milus, Oost and Holleman (2006), see document “Draaiboek expert panels” in folder 02 Research material.</p> <p>The advices and recommendations were written on post its, which are digitalized in document “ALLE verzamelde strategieën” in folder 03 Raw Database.</p> <p>The post its are, currently, stored in the closet of R.R. de Leeuw, room 128. Will be stored in the archive room of GMW.</p>
Description of the processed data <i>Describe how the processed data looks like.</i>	<p>In the folder 03 Raw Database:</p> <ul style="list-style-type: none"> - Raw digitalized data in Excel sheet, named “ALLE verzamelde strategieën” - Cleaned file in ATLAS.ti 7 file, named “Expert Panels Strategieën” and Excel sheet named “Analyse Expert Panels Strategieën” - Co-occurrence analyses from ATLAS.ti 7 exported in RTF, named “co-occurrence in atlas ti” - Folder named “ingedikt” with the comprised version of strategies (part of coding phase) <p>In the folder 04 Coding (phases) the steps of the coding phases are documented per phase, including data files in Excel sheet, Word and ATLAS.ti 7.</p> <p>This folder has a separate Dutch read me file describing the coding phases, named “Achtergrond Expert Panels en stappen in de strategieën”</p>
Data storage	
Stores raw database	<ul style="list-style-type: none"> - Raw digitalized data in Excel sheet, named “ALLE verzamelde strategieën” - The post its are, currently, stored in the

	closet of R.R. de Leeuw, room 128. Will be stored in the archive room of GMW.
Storages syntax and computer scripts/ description of coding phases/process	See folder 04 Coding (phases) on the Y-drive and read me file “Achtergrond Expert Panels en stappen in de strategieën” describing coding process. Same documents are on the X-drive of R.R. de Leeuw
Data archiving	
<i>Describe until when the data should be stored and from whom this date is set.</i>	Additional archiving requirements/agreements: Data will be stored until: 10 years after finishing complete PhD period; estimated 2030
Data ownership	University of Groningen
Data documentation	
Folder ‘Ethics committee application & approval’	Folder “General project approval” containing contract and document with approval of ethics committee
Folder ‘Research material’	<ul style="list-style-type: none"> - Draaiboek expert panels (Word) - Opdracht beschrijving casus (Word)
Folder ‘Raw database’	Folder “ruwe data”; “ingedikte strategieën” Documents: <ul style="list-style-type: none"> - 44 ingedikte strategieën (Word) - ALLE verzamelde strategieën (Excel) - Expert panel Achtergrond gegevens (SPSS) - Kappa berekenen (Excel)
Folder ‘Syntax, scripts, statistical logbooks’ → Coding (phases)	<ul style="list-style-type: none"> - Read me first file Expert panel and coding (Word) - Gecombineerde codebook (Word) - Kappa berekenen (Excel)
Folder ‘Figures and Tables’	<ul style="list-style-type: none"> - Figure 1 (TIF file) - Table 1 (Word) - Table 2 (Word) - Table 3a (Word) - Table 3b (Word) - Table 4 (Word)
Folder ‘Final publication’	<ul style="list-style-type: none"> - EJSNE Teacher Strategies to Support the Social Participation of Students with SEBD in the Regular Classroom (Pdf) - Teacher strategies to support xxxxxxxx – Manuscript (Word)
Others, loose files/documents	<i>Document: Read me first_DMP PUBDeLeeuwDeBoerBijstraMinnaert2017Word</i>

4.2 Example of a quantitative publication package

This example of a qualitative publication package is from a sub-study of a single PhD project. The image below is a screenshot of the folders within the publication package and contains raw data, final data set, methodology descriptions and process, figures and tables published in the article, the Word version and Pdf version of the article.



Administration	
Name	Dorinde Jansma
P-number	P255050
Name of research group	GION
Name of project	This is wrong, right? The role of Moral Components in Anti- and Prosocial Behaviour in Primary Education
Description of the study	The aim of this project was to obtain new insights into the relative contribution of four components of moral functioning to anti- and prosocial behaviour in primary education.
Funding agency	NWO Research Talent Grant 406-12-109
Institute project code	170514273
Project duration	PhD project July 2013 – August 2017
DMP version	Nieuwenhuis Institute Manual 2017
Project data managers <i>For every data manager: Name, contribution/description what is done regarding data collection</i>	n.a.
External collaboration	n.a.
Data collection	

Permits & licenses <i>Describe which type of informed consent was given, by whom, how and where consent forms are stored.</i>	<p>Approval for this research was granted by the ethical committee <i>Pedagogische Wetenschappen en Onderwijskunde</i>, at 21 October 2013. (See document in folder 01 Ethics committee application & approval)</p> <p>For all participants, informed consent was obtained from their parents.</p>
Description of the raw data <i>Describe what kind of raw data is collected</i>	<p>Data was collected regarding the four moral components and pro- and antisocial behaviour during four occasions. Both interviews and online questionnaires were used. The raw interview data are stored in the closet of D. J. Jansma, room 222, Grote Rozenstraat 3. Will be stored in the archive room of GMW. The raw questionnaire data is stored in Qualtrics.</p>
Description of the processed data <i>Describe how the processed data looks like.</i>	<p>In the folder c. Raw Database the raw digitalized data is stored in SPSS files. There is one file per measurement occasion.</p>
Data storage	
Stores raw database	<p>In the folder c. Raw Database the raw digitalized data is stored in SPSS files. There is one file per measurement occasion..</p>
Stores syntax and computer scripts/ description of coding phases/process	<p>In folder e. Statistical logbooks on the Y-drive there are syntaxes per study.</p>
Data archiving	
<i>Describe until when the data should be stored and from whom this date is set.</i>	<p>Additional archiving requirements/agreements:</p> <p>Data will be stored until: 10 years after finishing complete PhD period; estimated 2028</p>
Data ownership	
University of Groningen	
Data documentation	
Folder 'Ethics committee application & approval'	Folder b. b. Ethics committee application and approval containing contract and document with approval of ethics committee
Folder 'Research material'	<ul style="list-style-type: none"> - Coding schemes - Interview - Online questionnaire - Teacher questionnaire - Training and protocol
Others, loose files/documents	<i>n.a.</i>

Appendix 1. Checklist for publication packages

CHECKLIST PUBLICATION PACKAGES

Name:	P-number:	Package:	Date:
-------	-----------	----------	-------

<p>General</p> <ul style="list-style-type: none"> <input type="checkbox"/> Co-authors/supervisors access <input type="checkbox"/> Folder in right place <input type="checkbox"/> Name clear <input type="checkbox"/> Clear structure <input type="checkbox"/> Read me file <input type="checkbox"/> Publication (+ publication in Pure)
<p>Understanding study</p> <ul style="list-style-type: none"> <input type="checkbox"/> Description of how and by whom the data were collected or created <input type="checkbox"/> Questionnaire <input type="checkbox"/> Codebook <input type="checkbox"/> Sample/participants description <input type="checkbox"/> Ethics approval <input type="checkbox"/> Contributions per author
<p>Replicating results</p> <ul style="list-style-type: none"> <input type="checkbox"/> Digital data (raw) <input type="checkbox"/> Digital data (analyzed) <input type="checkbox"/> Syntax cleaning <input type="checkbox"/> Syntax analyses
<p>Future data</p> <ul style="list-style-type: none"> <input type="checkbox"/> Questionnaires in GMW archive <input type="checkbox"/> Digital data archived

Appendix 2. Sustainable data formats

Text processing

Preferred formats

- pdf/A
- Open Document Format: .ODT

Accepted formats

- RTF
- Postscript
- Word

Spreadsheets

Preferred formats

- Open Document Format: .ODS
- pdf/A

Accepted formats

- MS Excel
- Lotus 1-2-3
- Quattro Pro

Data files

Preferred formats

- SPSS portable files
- SAS transport files
- STATA transport files

Pictures

Preferred formats

- Tiff
- JPG (100% quality)

Databases

Preferred format

- ANSI SQL

Accepted formats

- MS Access 95 and higher
- dBase III+ en hoger
- DataPerfect
- Filemaker Pro 3.0 and higher
- Paradox

Plain text

Preferred format

- UNICODE with Byte Order Mark (UTF-8, UTF-16 of UTF-32)

Accepted formats

- ASCII (7 bit)
- ISO 8859 character sets
- MS-DOS codepages
- MS-Windows codepages
- Apple codepages

Mark-up*Accepted formats*

- XML (including XHTML) with DTD or scheme
- SGML (including HTML) with DTD

Audio / video*Preferred formats*

- Audio: MP3 (256 kbps), WAV, AIFF
- Video: MPEG-2, MPEG-4 AVC/H264, lossless AVI, QuickTime DV

Appendix 3. Read me first metadata file template

Research data management protocol

Read me first metadata file folder [insert folder name here]

Administration	
Name	
P-number	
Name of research group	
Name of project	
Description of the study	
Funding agency	
Institute project code	
Project duration	
DMP version	
Project data managers <i>For every data manager: Name, contribution/description what is done regarding data collection</i>	
External collaboration	
Data collection	
Permits & licenses <i>Describe which type of informed consent was given, by whom, how and where consent forms are stored.</i>	<p>Approval for this research was granted by [fill in what is applicable]. <i>Remove if not applicable:</i> For all participants, active informed consent was obtained from parents or legal representatives/teacher(s)/school director/care taker/institution/ employee of [name of institution].</p> <p>For all participants, passive informed consent was obtained from parents or legal representatives/teacher(s)/school director/care taker/institution/ employee of [name of institution].</p>
Description of the raw data <i>Describe what kind of raw data is collected</i> <i>If possible how many per participants, including number of participants</i>	
Description of the processed data <i>Describe how the processed data looks like. For example:</i> <i><u>Excel worksheet</u></i> <i><u>TIF File</u></i> <i><u>R Studio 1.0.44 file</u></i> <i><u>SPSS Statistics Data Document</u></i> <i><u>SPSS Statistics Syntax File</u></i>	
Data storage	
Storages raw database	

<p>Storages syntax and computer scripts/ description of coding phases/process</p> <p><i>Describe where syntaxes and computer scripts are stored. For example: <u>X-drive, Y-drive, external hard disk including owner</u></i></p>	
<p>Data archiving</p> <p><i>Describe until when the data should be stored and from whom this date is set. For example:</i></p> <ul style="list-style-type: none"> - <i>If a journal requests data storage name the journal and publication date</i> - <i>Data storage rules from University of Groningen, note down the expiration date</i> 	<p>Additional archiving requirements/agreements:</p> <p>Required by/ agreed to: [<u>fill in institution/journal</u>]</p> <p>Data will be stored until: [<u>fill in expiration date date</u>]</p> <p><i>In addition provide the details of requirements/agreements.</i></p>
<p>Data ownership</p>	<p>University of Groningen</p>
<p>Data documentation</p>	
<p>Folder 'Figures'</p>	<p><i>Note down the names of the files/documents in the folder 'Figures'</i></p>
<p>Folder 'Research material'</p>	<p><i>Note down the names of the files/documents in the folder 'Research material'</i></p>
<p>Folder 'Syntax, scripts, statistical logbooks' or 'Coding process'</p>	<p><i>Note down the names of the files/documents in the folder 'Syntax, scripts, statistical logbooks'</i></p>
<p>Folder 'Final publication'</p>	<p><i>Note down the names of the files/documents in the folder 'Final publication' (mostly pdf)</i></p>
<p>Folder 'Ethics committee application & approval'</p>	<p><i>Note down the names of the files/documents in the folder 'Ethics committee application & approval'</i></p>
<p>Folder 'Raw database'</p>	<p><i>Note down the names of the files/documents in the folder 'Raw database'</i></p>
<p>Folder 'Cited literature'</p>	<p><i>In this file you provide all files/documents of the cited literature.</i></p> <p><i>When you have books, websites and others list these items in a Word document and refer where these items can be found</i></p>
<p>Others, loose files/documents</p>	<p><i>Note down the names of the files/documents that are not filed in the required folders</i></p>