Publication Packages
&
Research Packages
Content

1 Publication and research packages: A good scientific practice 3
2 Who, what, where, and when? 4
3 Publication packages 6
4 Example of a publication package 8
5 Research packages 12
6 Example of a research package 14
7 Appendix 1. Checklist publication packages 16
8 Appendix 2. Sustainable data formats 17
1 Publication and research packages: A good scientific practice

Following the guidelines of the University of Groningen¹ and the Association of Universities in the Netherlands (VSNU)², the Department of Sociology recently implemented policy to stimulate researchers to handle their research data in a responsible way.

In short, researchers in the department are from September 2016 onwards requested to document important steps that are taken during their research and store their digital data in special folders or packages. In the Department of Sociology there are two types of packages:

- **Publication packages** contain all data, syntaxes, and other important information needed to replicate results that are published in a scientific journal or book (chapter)

- **Research packages** contain all data, syntaxes, and other important information belonging to a larger research project (i.e., projects where several publications are based on one dataset such as the KiVa project)

The most important reason to document your research and store your data in a safe place is that you and your co-authors/research colleagues/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** and **replicate** your research.

---

2 Who, what, where, and when?

Who?
All members of the Department of Sociology (staff, postdocs, and PhD students) are responsible for ensuring that there is a publication package for every accepted article in a scientific journal or book (chapter) of which they are the first author. All principal investigators of the Department of Sociology are responsible for ensuring that there is a research package for their research projects.

Consistent with the oath at the PhD defense, the policy 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). The graduate school will not send out the thesis to the reading committee before the packages are in place. Please send an e-mail to the coordinator\(^3\) when you have completed the packages. The coordinator will check whether all necessary files are there\(^4\) and inform the graduate school.

By default, only the owner of the package has reading and writing rights. The coordinator and technical support director have reading rights.

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

https://www.formdesk.com/rug/GMW_Request_access_Sociologie_folder

Access rights need to be requested per package. At present it is not possible for co-authors/research colleagues/supervisors at other universities to access the packages, but this should be possible in the future.

What?
The packages contain all (quantitative and qualitative) data used by members of the Department of Sociology. Note that with ‘data’ we mean (primary and secondary) empirical data but also other types of data, such as meta-analyses. When for certain reasons it is not possible to include the data in a package (ethical guidelines, insurmountable logistical problems, privacy-related reasons or formal property rights), contact the coordinator.

The packages are in principle intended to verify and replicate research. They are not meant to share data for reuse. However, the policy of the Department of Sociology is that data become publicly available unless there are ethical, juridical, contractual reasons not to do so. Sharing data contributes to the scientific body of knowledge and safes effort, time, and costs. Grant suppliers (such as NWO, KNAW, and ZonMw) and journals increasingly demand that

---

\(^3\) Currently the coordinator is Beau Oldenburg. Beau Oldenburg is a fourth-year PhD student who has done an internship at DANS and followed several courses related to data management (e.g., GESIS and Research Data Netherlands).

\(^4\) At this moment the coordinator will only check whether all necessary files are in place. See the checklist in Appendix 1. In the near future, the faculty’s integrity committee will possibly check the packages more in-depth (e.g., rerun analyses).
research data are made accessible for other researchers in so-called trusted repositories (e.g., DANS).

**Where?**
Every member of the Department of Sociology has access to the safe digital environment of the university: the Y drive. On the Y drive a folder has been created for the publication and research packages (Y:\Staff\gmw\sociologie)\(^5\). 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., WALM). Contact the coordinator when you cannot find the folder with your P-number (for instance, when you recently started working at the department). Also contact the coordinator when you want to make a new research package.

All data on the Y drive are retained for at least ten years. The Y drive uses a backup system that makes sure files cannot be deleted (everything stored on that drive can be retrieved, even if someone edits or deletes data after it has been saved). When you leave the department you can request to keep you P-number so that you will still be able to access your package.

**When?**
It is wise to make the package as soon as possible after the completion of the research project. In any case, the package should be ready within three months after the publication or completion of the research project.

\(^5\) The RUG is currently developing an alternative for the Y drive. This alternative infrastructure can hopefully be accessed by researchers who have no (longer a) P-number.
3 Publication packages

Publication packages are made for every published scientific article or book (chapter) of which a department member is a **first author**. PhD students make a separate folder for each chapter of their dissertation (excluding the introduction and conclusion).

On the Y drive a folder has been created for the publication packages (Y:\Staff\gmw\sociologie). 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., WALM). 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. PUBJansenBakkerSmit2015). In the case of a PhD thesis please clearly indicate the chapter of the thesis in the name.

Given the different types of research in the department it is difficult to provide an exhaustive list of information that should be included. A guideline is that a colleague who was not involved in the study should be able to understand your study and replicate your results based on the information in the package. For most studies the package would at least contain the following information (see also the checklist in Appendix 1):

- A **readme** file explaining which files are included in the package
- The published article or book (chapter)

**Information that is needed for others to understand your study:**
- Description of how and by whom the data were collected or created
- Overview of questions that were asked (i.e., questionnaire)
- Explanation of the variables (i.e., codebook)
- Description of sample/participants
- Information on whether your study was evaluated by an ethics committee (if so, include application and letter of approval)
- A description of the contributions made by each author/co-author, for example ‘devising and organizing the project’, ‘data collection’, ‘data analysis’, ‘critical reading of the text/feedback’, and ‘writing of the article/PhD thesis’

**Information that is needed for others to replicate your results:**
- Raw (or original) digital data. Contact the coordinator when it is not possible to store the data in the publication package (ethical guidelines, insurmountable logistical problems, privacy-related reasons or formal property rights)
- In the case of non-digital data state where the data are stored and how they can be accessed
- 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

---

6 We strongly recommend storing the data in a sustainable format (see the preferred formats in Appendix 2). Data in sustainable formats can still be accessed when for instance certain software is no longer available.
- Syntax for created variables/scales and analyses or coding schemes for qualitative data
- Other digital relevant research material: stimuli, instructional texts, experiment leader protocols, video material, software for simulation studies, ...

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

https://www.formdesk.com/rug/GMW_Request_access_Sociologie_folder

The package should be ready within three months after the article or book(chapter) is published. In line with the university regulations, PhD’s can only defend their thesis once the publication packages for all chapters (excluding introduction and discussion) are approved. If staff members fail to make a publication package, they can be sanctioned (e.g., reduction of research time).

Please contact the coordinator when a publication package is completed. Do not forget to also add your article to Pure!
4 Example of a publication package

1. **Codebook**: The codebook contains information on the variables in the data. In this codebook the name of the variables, a description of the variables, the source of the information (e.g., teachers or students), the value labels, original scales and references to the original scales are included.
2. **Data archiving form:** This document contains information on whether the project was reviewed by an ethical committee. In addition, the contributions of each author are described. Furthermore, is described who collected the data. Finally, in this file is described that the data are stored open access at data archive DANS and that the paper questionnaires are archived at the GMW archive (and will be kept there for 10 years).

3. **Description of the data:** Because this was a rather complex data-collection (combining data from two sources) I made an extra file with an in-depth explanation of the data.

**EXPLANATION OF ‘KIVA TEACHER DATA’**

The following sample description is an adaption of the sample description of Oldenburg et al. (2015). Please refer to this description in the following way:


Note that Oldenburg and colleagues did not use the full dataset as provided here but used a subsample.

In dataset ‘Kiva teacher data’ data from two different sources are combined:
- Dutch primary school students
- Teachers of these students

**Dutch primary school students**
The data are collected amongst students and teachers who were part of the evaluation of the Dutch version of the KiVa anti-bullying program. The KiVa program is developed in Finland (e.g., Kiirni et al., 2011) and aims to prevent and reduce bullying in elementary schools. KiVa is currently being implemented and tested in several countries, including the Netherlands. The school year in the Netherlands ranges from the end of August to the beginning of July. In the fall of 2011 all 6,966 regular Dutch elementary schools (Statistics Netherlands, 2012) received an invitation to participate in the KiVa anti-bullying program. Special elementary schools and schools for children with special educational needs could not participate in the KiVa program and were hence not invited to participate.

The 64 schools that were willing to participate were randomly assigned to either the control

4. **Data analyses, KiVa teacher data cleaned and linked, and KiVa teacher data not linked and not cleaned:** I included the raw data and the data on which the publication was based in the publication package. It is important to always keep the original data file. This way the results can always be reconstructed.

5. **KiVaNL_Oldenburg14:** Published article.
6. **Questionnaire**: I included the questionnaire as presented to the respondents in the publication package.

7. **Syntax analyses & syntax data cleaning**: I cleaned the data using SPSS and analyzed the data using Stata.
8. **Syntax Marijtje**: Syntax used for checking outliers, making plots, etc.

9. **Outliers**: Overview which outliers were eventually deleted.

**Readme file**: A file explaining which files are in the package.
5 Research packages

Research packages contain all data, syntaxes, and other important information belonging to a larger research (i.e., projects where several publications are based on one dataset such as the KiVa project). All principal investigators of the Department of Sociology are responsible for ensuring that there is a research package for their research projects.

On the Y drive a folder has been created for the research packages (Y:\Staff\gmw\sociologie). Contact the coordinator when you want to make a new research package.

Given the different types of research in this department it is difficult to provide an exhaustive list of information that should be included. A guideline is that a colleague who was not involved in the project should be able to understand your research. For most research projects the package would at least contain the following information:

- A readme file explaining which files are included in the package
- The published articles or books (chapters) within the project
- A description of the contribution made by each researcher. Types of contribution include ‘devising the project’, ‘organizing the project’ and ‘data collection’
- Information on who owns the data
- Overview of questions that were asked (i.e., questionnaire)
- Explanation of the variables (i.e., codebook)
- Description of sample/participants
- A description of whether the project was evaluated by an ethics committee (if so, include application and letter of approval)
- Raw (or original) digital data. Contact the coordinator when it is not possible to store the data in the research package (e.g., due to ethical guidelines, insurmountable logistical problems, privacy-related reasons or formal property rights)
- In the case of non-digital data state where the data are stored and how they can be accessed
- 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: stimuli, instructional texts, experiment leader protocols, video material, software for simulation studies, ...

---

We strongly recommend storing the data in a sustainable way (see the preferred formats in Appendix 2). Data in these formats can still be accessed when for instance certain software is no longer available.
It is important that others involved in the research (co-authors/supervisors) can access the package. Access can be requested via this form:

https://www.formdesk.com/rug/GMW_Request_access_Sociologie_folder

The package should be ready within three months after the research project is completed. If owing to unforeseen circumstances data collection has not yet ended by that date contact the coordinator, the deadline may be extended. If principal investigators fail to make a research package, they can be sanctioned (e.g., reduction of research time).

Please contact the coordinator when a research package is completed.
6 Example of a research package

The research package of the KiVa project is used as an example. The KiVa package can be found in the following location:

The package contains the following folders:

1. **Documentation**: We have combined the KiVa research data management plan and the KiVa codebook in one document so that anyone who was not directly involved in the KiVa project can find all necessary information in one document. The documentation contains among other things information on how the data are handled, who owns the data, a description of the contribution made by each researcher, a description of the sample, and information on the variables. Please contact b.oldenburg@rug.nl or d.r.veenstra@rug.nl if you would like to use this documentation as an example.
2. **Questionnaires:** This folder contains screenshots of the online questionnaires as presented to the respondents.

3. **Data:** This folder contains the cleaned data. The raw data could not be included in the package, but we describe where these data can be found.

4. **Syntax:** This folder contains all syntax that are used to clean the different waves.

5. **Publications:** We included the KiVa publications in this package.

6. **Readme file:** A file explaining which files are in the package.
## Appendix 1. Checklist publication packages

### CHECKLIST PUBLICATION PACKAGES

<table>
<thead>
<tr>
<th>Name:</th>
<th>P-number:</th>
<th>Package:</th>
<th>Date:</th>
</tr>
</thead>
</table>

### General
- □ Co-authors/supervisors access
- □ Folder in right place
- □ Name clear
- □ Clear structure
- □ Read me file
- □ Publication (+ publication in Pure)

### Understanding study
- □ Description of how and by whom the data were collected or created
- □ Questionnaire
- □ Codebook
- □ Sample/participants description
- □ Ethics approval
- □ Contributions per author

### Replicating results
- □ Digital data (raw)
- □ Digital data (analyzed)
- □ Syntax cleaning
- □ Syntax analyses

### Future data
- □ Questionnaires in GMW archive
- □ Digital data archived
8 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