

Research Data Management Protocol Faculty BSS – Implementation Guide

December 2022

1 Introduction

This document is the implementation guide for the data management protocol for the faculty of BSS (data-management-protocol-bss-policy). In this document, the current storage, archiving and open science implementation is described.

2 What to store, and where to store it?

For an extensive description of what needs to be stored see the policy document.

Table 2.1 provides an overview of the most commonly used folders and their content. As stated before, different types of content may be required, depending on the nature of the study and the data. Thus, for every study that is conducted, proper documentation regarding the data collection process, ethics approval, research materials, data files, and data processing/analysis is required. If the study was published, the publication should also be included in the package.

Table 2.1. Most commonly used folders and description of content

| <i>Folder/file</i> | <i>Contents (when applicable)</i> |
|-------------------------------|--|
| Data archiving form | <ul style="list-style-type: none">• A data archiving form specifying all those involved in the study and their respective roles (e.g., Principal Investigator, co-researcher, research assistant, etc.), the period and location of data collection, the location of non-digital materials, etc. |
| Ethics application & approval | <ul style="list-style-type: none">• A pdf file of the ethics application and ethics approval |
| Research materials | <ul style="list-style-type: none">• The survey or experimental program(s) used• Explanation of the variables (i.e., a codebook)• Media files presented to participants (or stable links to these files)• Other relevant materials relating to the study (e.g., experiment protocols) |
| Data and syntax | <ul style="list-style-type: none">• Raw (or original) digital data• Syntax and/or statistical logbooks of the processing of the raw data |

| | |
|-------------|---|
| | <ul style="list-style-type: none"> • Other relevant materials relating to data processing, such as coding schemes for qualitative data • A properly documented and edited data file + syntax for the actual statistical analyses |
| Publication | <ul style="list-style-type: none"> • The published articles or books (chapters), if applicable |
| Readme file | <ul style="list-style-type: none"> • A <i>readme</i> file describing which files are in the package (i.e., an index) |

2.1 Folder structure of a single publication data package

The data package should be well *organized*. By using clear **file** and **folder names** (as well as subfolders for the different categories of files) and including clear and comprehensible **comments** in syntax and other files, researchers can make sure that everything is clear even if materials are accessed years after the study has been done. Providing extra ‘Read me’ documents with meta-information is also a useful and sometimes necessary tool here.

A well-organized folder could look something like the one displayed in Figure 2.1. Different components of the study are stored in separate and clearly named folders. Note that researchers do not have to use *exactly* this structure or *exactly* these names, but something along these lines probably makes sense and will be easy to use.

Figure 2.1. Example of the structure of a data package for a published article



2.2 Folder structure for larger research projects (containing sub-packages)

The data package in Figure 2.1 contains all information relevant to a study resulting in a single publication. However, more often than not publications are not singular entities but part of a larger research project. The easiest way to organize such materials is to structure them in project folders. This structure could be used for a ‘formal’ project, such as a Vidi grant project, or for more ‘informal’ projects – basically, a research theme pursued by a researcher.

Take, for example, the fictitious project “For the Win”, in which a professor has successfully acquired a grant for a five-year research project with two PhD students and a Postdoc. This means that, by the end of the project, two dissertations and several separate studies will be present in the project folder. Each dissertation should be placed in its own subfolder; the same applies to the other separate studies. Datasets that are used for multiple publications only need

to be included in the project folder once, as long as the subfolder for each individual publication contains information about where the dataset is located.

The subfolders and documentation in the packages should match the complexity of the project. It is important that both the superordinate project level and the subordinate levels of the different PhD and postdoc projects are properly organized and documented. At the superordinate level, relevant information to store would be, for example, the project proposal (grant application), memos regarding the project as a whole, etc. At the subordinate levels, actual storage of research materials and data become relevant; this is where the individual data packages would be. If multiple subordinate projects make use of the same data, this should be clearly documented at the superordinate (project) level.

Many other studies are not part of a formal (grant-based) project, but follow a particular research theme and can therefore still be considered part of a 'project'. For example, a researcher might be particularly interested in collective emotions during group tasks. Over the years, this researcher may perform many studies on this topic, which might be best grouped together in a single superordinate folder 'Group Emotions' or something similar.

Important note: It is up to researchers to create and maintain a folder structure that works for them, as long as transparency and systematic storage are safeguarded.

2.3 Storing files

The university currently doesn't have a storage facility that can satisfy all the requirements of the research data policy. In the future, the RDMC system (currently under development) may become the default location for research data. For now, the best and safest storage option the university has is the Y-drive. The packages should be stored on the Y drive. The Y drive is secure, which makes it an appropriate place to store most data. Moreover, the Y drive is routinely backed up, which means that all files stored there are safe from accidental erasure or file damage. The Y drive does *not* allow for sharing data with external researchers or publishing data to open science platforms. In the future, a new environment will be selected that does make these things possible and satisfies all the requirements described in the research data policy of the faculty.

Every staff member of the faculty has a personal folder on Y. Approach your department's contact person listed in section 1.3 if you cannot find your folder or when you are new to the University of Groningen. If you do not use a RUG workstation, the Y drive can easily be accessed using the Citrix Receiver or by going to <http://uwp.rug.nl/>.

All data on the Y drive are retained for at least ten years.³ 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).

Note that the Y drive is intended to let one or more employees from the University of Groningen access a folder. For cooperating with others from outside the university the use of Unishare is highly recommended. You can access Unishare by going to <https://unishare.nl/>.

The X drive can only be accessed by you and is a good place to store data at an early stage of your research project, when you did not make a data package yet.

2.4 GDPR and privacy

Some data should either not be stored, or should be stored differently than ‘normal’ data. This is usually the case when there are privacy issues. Especially for research dealing with sensitive

³This period follows the recommendation of the *Richtlijn archivering wetenschappelijk onderzoek voor Nederlandse faculteiten Maatschappij-en Gedragwetenschappen* from 2017.

personal data, researchers are strongly urged to contact Research Support BSS at an early stage for assistance with risk assessment of the research data, pseudonimization techniques, legal review, and data security measures.

As a rule, consistent with national guidelines for data archiving, the data package should *never* contain data that are not anonymized or pseudonimized.⁴ Such data should be considered 'personal data', and includes (but is not limited to): names, birthdays, (email) addresses, location data (GPS), family status or details, and everything else that can be used to retrace someone's identity. In these cases, the data package should contain the first version of the data that *can* be anonymized or pseudonimized (for example, a transcript using pseudonyms in the case of audio recordings). The package will have to include information specifying where and how (and to whom) these original data are available.

If you need to store personal data in which participants are easy to identify (e.g., audio, video), it should be stored in a folder on the Y drive, separately from the data package. A folder on the Y drive can be requested via this link:

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

Make sure to request a folder for each research project, separately. This way you minimize the number of people that (unnecessarily) have access to the data. The requested folder should be located on the Y drive of the faculty in the folder of your department (e.g. y:\staff\gmw\psychologie).

If you are dealing with sensitive personal data, it is recommendable to contact Research Support BSS for advice on data management.

⁴The difference between pseudonymized and anonymized data is the following: anonymization irreversibly destroys any way of identifying the data subject, while pseudonymization substitutes the identity of the data subject in such a way that additional information is required to re-identify the data subject