



Research Data Management

Research Institute of Child Development and Education
2022

TABLE OF CONTENTS

RESEARCH DATA MANAGEMENT	2
RICDE DATA MANAGEMENT POLICY	3
WHAT IS A DATA PACKAGE?	3
WHEN SHOULD A DATA PACKAGE BE SHARED?.....	3
WHICH DATA PACKAGES MUST BE SHARED?.....	4
WHICH DATA PACKAGES CAN BE SHARED ADDITIONALLY?	5
HOW TO SHARE A DATA PACKAGE?.....	5
DATA MANAGEMENT EXTERNAL PHD STUDENT.....	6

Research Data Management

Research data management (RDM) serves three purposes: (1) findability; RICDE must be in possession of all data collected under its responsibility; (2) traceability; data must be saved with complete provenance, to guard scientific integrity and as a back-up, so that researchers (or their successors) are always able to check, redo and continue their work; (3) data sharing; data should be made available to the community, to the extent possible, so that other researchers can use the data to answer new research questions.

Good RDM is also in the interest of the researchers themselves. Storing data properly and backing-up regularly prevents data loss, allowing researchers to always find their data, ensuring the quality of their scientific practice. Moreover, it stimulates collaboration with others, who will find it easier to understand and re-use data, which makes RDM cost and time efficient.

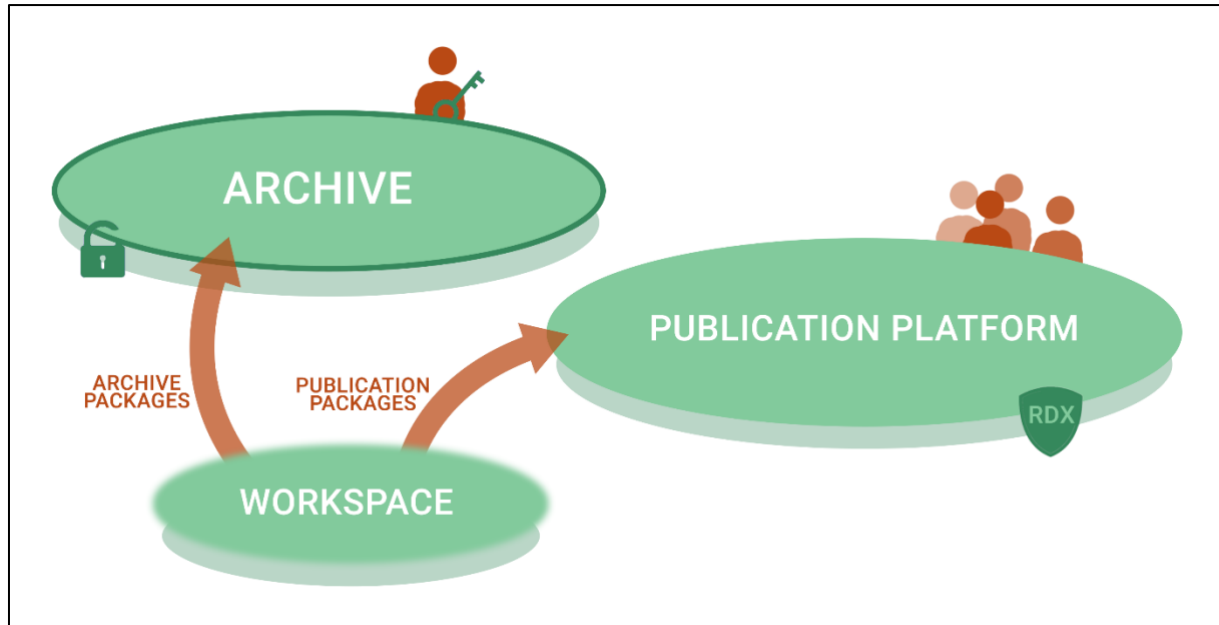
In view of these purposes, we want our data to be FAIR: Findable, Accessible, Interoperable, and Reusable¹. However, data sharing is subject to conditions imposed by laws and regulations (such as the General Data Protection Regulation; GDPR), and to the interests of the original researchers and the university. We therefore distinguish between FAIR archiving (closed) and FAIR publishing (open).

Important: We strive to *archive all data* with full provenance (FAIR for the institute), to the extent possible, subject to laws and regulations, contracts and agreements, and informed consent. We also strive to *publish a selection* of suitable data (FAIR for the world), subject to the same conditions as to the archival of data.

In this document we distinguish between workspace storage, archiving, and publication. Figure 1 explains this distinction.

¹ Wilkinson, M. D., et al. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, [160018]. <https://doi.org/10.1038/sdata.2016.18>

Figure 1: Workspace storage, archiving, and publication of data in separate locations.



Note: Temporary storage space is provided as WORKSPACE storage, for processing data, for the duration of the research project. Immediately after data collection, and at any other time when warranted, the researcher will submit data packages for archiving. The ARCHIVE stores data packages with full provenance for an indefinite period of time, closed, only accessible to the data steward, after permission from the research director. Data packages that are suited for publication can be submitted to the PUBLICATION PLATFORM, which holds data packages open, for an indefinite period of time.

RICDE Data Management Policy

The Data Management Policy of the Research Institute Child Development and Education (RICDE), requires researchers to share so-called data packages with the data steward. What a data package is, and what it should contain is explained under “[What is a Data Package?](#)”. These data packages are then copied by the data steward to the closed archive of RICDE on the FMG storage. The data will be archived indefinitely, at a secure location (GDPR compliant), that is only accessible to the data steward, after permission from the research director.

What is a Data Package?

A data package always contains one or more data sets and other relevant files. What those other relevant files are depends on the purpose for which the data package was compiled. We distinguish between:

1. Data packages for archiving (with full provenance, and (thus) closed)
2. Data packages for publication (open, with use conditions)
3. Data packages for transfer (with transfer agreement)

When should a data package be shared?

During a research project, multiple data packages will be compiled. For example, once immediately after data collection, once each time a manuscript is submitted for publication, and perhaps once more if the researcher leaves the university but wants to take some of the

data with him or her. Below are the data packages that must be shared with the data steward during each project (data packages for archiving). This is followed by examples of optional data packages (data packages for publication and/or for transfer).

Which data packages must be shared?

Immediately after the data collection has been completed, or a substantial part, the first *data package for archiving* is compiled.

This contains:

- A. The original data (raw, possibly sensitive).
- B. A description of the dataset (e.g., a research protocol with operationalizations and methods of sampling and data collection),
- C. A codebook that lists all the names of the variables and their meanings, including all possible values that these variables can take and meaning of all those values.

If identifying information is no longer needed, then the data should be de-identified immediately after archiving (through anonymization or pseudonymization), so that the original, personal data can be removed from the workspace (e.g., OneDrive) that is used for data processing and analysis. The de-identified version of the data should be archived as well.

After a substantial part of the analysis has been done (e.g., when the data analysis for a manuscript has been completed), another *data package for archiving* is compiled.

This contains:

- A. The de-identified version of the data (anonymous or pseudonymous),
- B. The processed dataset,
- C. A description of the dataset (for example, a research protocol, a scientific article or other document in which all operationalizations of all constructs are described, and the method of data collection, etc.),
- D. A codebook that lists all the names of the variables and their meanings, including all possible values that these variables can take and meaning of all those values,
- E. A description of full provenance, possibly including reference to the first data package for archiving, and (a description or) the (annotated) scripts that need to be executed to get from the original data to the processed dataset.

Researchers can of course add more information to the data package or share more information with the data steward in addition to the above required documents. This may include the names of all researchers involved in the research project with a description of their roles, a description of deviations from the study procedure compared to the original protocol, permission from the Ethics Committee, correspondence with the grant provider, etc.

Important: Some data packages can be created multiple times during a project. For example, during a longitudinal study, or when a dataset has to be processed in different ways to make it suitable for statistical analyses.

Which data packages can be shared additionally?

When a researcher wants to publish his or her data, a *data package for publication* is compiled.

This is similar to the data package for archiving, but:

- A. The provenance does not have to be complete, or can even be omitted completely (the raw data and intermediate steps to the final data do not have to be published, or can't be published, because of privacy),
 - a. For guidance with data publication, you can always contact RICDE's data steward: datasteward-pow-fmg@uva.nl

When a researcher wants to take data with him or her when he or she leaves the university, a *data package for transfer* is compiled.

This contains:

- A. One or more data sets, preferably only de-identified versions of the data sets.
- B. A description of the dataset (for example, a research protocol or an article in which all operationalizations of all constructs are described),
- C. A codebook that lists all the names of the variables and their meanings, including all possible values that these variables can take and meaning of all those values,
- D. Data transfer agreement, signed by researcher and research director. If the data package contains sensitive data, then the transfer agreement must also stipulate privacy and security responsibilities. If applicable, please contact RICDE's data steward: datasteward-pow-fmg@uva.nl

How to share a data package?

If you have registered your research project in RMS (live since September 2022) you can upload your data package by following the instructions under the RMS module "data archiving".

If you have not registered your research project in RMS, for you have used the old system, you can follow these instructions:

1. Prepare a data package for archiving in your own workspace (e.g., Research Drive or UvA OneDrive).
2. [Compress](#) your data package before uploading, and make sure your data package has the following name structure:
 - a. *Surname_type of data package (archive, publication, transfer), e.g., Bakker_archive*

3. You can “drop” your data package for archiving by following [this link](#). After which the Data Steward will archive your data package in the closed RICDE archive.
4. Notify the data steward that you have dropped your data package, by sending an email to: datasteward-pow-fmg@uva.nl

Data Management External PhD student

When you are an external PhD student there are certain conditions under which you are not required to upload a data package with *full provenance*, but you only need to upload a smaller subset of the data package. This applies in the following situation:

1. Your data includes personal information,
2. You have stored, and archived your data (including personal information) at a third party,
3. There is no privacy agreement in place.

If these three conditions apply to you, you are asked to prepare a smaller version of a data package including all that you can share (anonymized data), a description of the data, a codebook, the manuscript, etc. You are also required to state where the actual data are stored (e.g., closed archive of the HvA), and how access can be obtained (e.g., contact information of the data manager at the HvA).

Instructions on how to share your data package can be found [here](#).