Guideline on Data Handling and Methods Reporting

Tilburg School of Social and Behavioral Sciences (TSB) Science Committee

Introduction

The scientific principles of Honesty, Scrupulousness, Transparency, Independence, and Responsibility in the [Netherlands Code of Conduct for Research Integrity](#) require that research data are managed responsibly. To improve data management, the deans of the Dutch Schools of Social and Behavioral Sciences (*Disciplineoverleg Sociale Wetenschappen*) agreed on a guideline for handling social sciences scientific research data in line with the Code of Conduct for Research Integrity and the General Data Protection Regulation (GDPR). This guideline for archiving scientific research formulates preconditions for the archiving of data, materials, and information that form the basis of publications to ensure transparency and reproducibility of results.

The Guideline on Data Handling and Method Reporting (DHMR) follows this national guideline. In this document, the national guideline has been applied at School level. It is explained to whom this Guideline applies, when a data package is needed, and what a data package should contain.

Implementation

This Guideline applies to all employees of the Tilburg School of Social and Behavioral Sciences (TSB) who carry out research in temporary or permanent employment, to all PhD candidates and to all Research Master's students. This Guideline also applies to endowed professors affiliated with the university. The Guideline does not apply to Bachelor's and one-year Master's students, unless the research of these students leads to a scientific publication. The research of Bachelor's and one-year Master's students is currently under the formal responsibility of their supervisors.

When do you need a data package?

Those to whom this Guideline applies must ensure that a data package is available in English for each empirical article, book, or chapter of which they are the first author. The Guideline applies to data collected by TSB staff and to statistical meta-analyses, qualitative research, data simulations, and analyses/secondary analyses of data collected by third parties (persons or bodies that are not TSB students or staff) or available in databases.

A data package needs to be stored within a month after the final publication of the manuscript. A Research Master's student needs to have his/her data package stored when submitting the Master's thesis. PhD candidates must store a data package for each empirical chapter of their thesis when submitting the manuscript to the reading committee. In the case of a monograph, one data package is sufficient. After the publication package has been saved, it is fixed and cannot be changed (read-only).

Responsibilities

If the first author of an article, book, or chapter is employed by TSB, he or she is always responsible for archiving the data package as described below. If a second, third, etc. author is employed by TSB, he or she must know how and where the data and data package is stored, especially when the first author does not work at the School. The role of each author must be specified in the metadata.

For PhD candidates and Research Master's students, the principal thesis supervisor or immediate supervisor is responsible for the storage of a data package. The principal thesis supervisor or immediate supervisor can delegate this task, but remains responsible for it.

Elements of the data package

A data package consists of nine elements: metadata and data collection, raw database, data storage, material, statistical processing, processed database, access and verification, retention period, and accepted manuscript. All elements are described below. A checklist of all the elements is included at the end of this document.

Metadata and data collection

The data package must include metadata providing information on when and by whom each file in the package was created and edited and who has/had access to the package. In the metadata, it must also be indicated whether an ethical review has been carried out, including the details (the ERB application number) of the review. The role of each author/co-author must also be clearly described (e.g., writing specific parts of the paper, analyses, etc.). In addition, the data package must include detailed information on who collected the data, at what location, and the date/period of time the data was collected. This also applies to data collected by third parties insofar as this information is available. In the case of data from existing databases and repositories, the original version and date of extraction must be specified. In the case of external finance/grants, it must be specified who provided these. The documentation of when the manuscript was accepted also needs to be included in the metadata, including the reference (reference with DOI is acceptable).

Note: Many data repositories and backup systems implement metadata including user access as standard. This is not always the case with the tracking of changes, and, therefore in some cases, a digital logbook will need to be created. This logbook can consist of, for example, an Excel or Word document in which this information is specified and tracked. The roles of authors/co-authors can be classified as follows: "devising and organizing the project", "data collection", "data analysis" and "article writing".

Information on data collection can be described in the publication and/or a file in the data package. In the case of data collected by TSB staff or students themselves, these descriptions are contained in lab journals providing detailed information on data collection and input. In the case of secondary data or data from other researchers or bodies, this information must be sufficiently specific in the form of contact data for the other researchers and clear descriptions of the sources from which and dates on which the researcher obtained/downloaded the data.

Raw database

The data package must contain the "raw database" as initially available to the TSB staff member digitally or in a digitalized form (e.g., scans of paper questionnaires). This must be a dated, read-only file. The raw database should be fixed upon completion of the data collection. The date of the fixation of the source data must be registered in the data package. This file is generally created when the data becomes available, i.e., before the article is written. Data in the raw database must be anonymized in accordance with ethical guidelines. The data package must *not* include information that identifies persons nor consent forms (which must be retained elsewhere, unless

other arrangements have been made on account of ethical objections). In the case of non-digital source materials (physiological findings, audio recordings, paper questionnaires, etc.), it must be stated where these are stored and how the data can be traced. In cases where the TSB researcher is not allowed to store the data on account of ethical guidelines, insurmountable logistical problems, privacy-related reasons, or formal property rights, the reasons for the absence of the data from the package must be clearly indicated. Researchers must strive for transparency in cases of such data management limitations (e.g., by means of a notification in the author's note to the article). In cases where the raw data files are stored and accessible in an external archive (such as DANS or DataverseNL), a source reference to the files in this archive is sufficient. This externally archived raw data can be primary or secondary data.

Note: The raw database is the first database obtained digitally by the TSB staff member. Ideally, this should be the first digital file created, but this file may also include datasets/secondary datasets from third parties or, for example, files downloaded from data repositories or other types of publicly available databases that have been used. Self-collected data includes data such as questionnaires or data collected by means of online surveys, computers, or measuring tools. In the case of paper questionnaires or documents, the researcher is expected to take responsibility for scanning or otherwise storing these paper questionnaires or documents. If ethical guidelines, insurmountable logistical problems, privacy-related reasons, or formal property rights make it impossible for the entire raw database to be included in the data package by the TSB staff member this must be made sufficiently clear and substantiated, with documentation, in a file in the data package. In such cases, the aim should be to include a random sample from the original database or a screenshot of part of the raw data in the data package.

Material

The data package must contain all the digital research material (or material for scanning) used in the research project that is needed to replicate the research, including questionnaires, stimuli, instructional texts, experiment leader protocols, video material, software for simulation studies, computer scripts, etc. Any references made to the source of material published elsewhere must be accurate and sufficiently specific.

Note: The principle of replicability applies here. In principle, a peer should be able to replicate the research based on the documents in the data package without the need for additional information.

Statistical processing

The data package must contain syntax, computer scripts, or statistical logbooks of the processing of the raw data so that analyses such as those reported in the article can be replicated from the raw database. The use of syntax, computer scripts or logbooks should in principle result in the processed database referred to below.

Note: The options here include not only those in SPSS and other statistical software to create the syntax automatically but also e.g. commands in R or other types of programs. If the statistical

software used makes it impossible to save commands automatically, a logbook should be kept specifying the analytical steps employed in an understandable manner.

Processed database

The data package must include a sufficiently documented and processed database enabling a peer to replicate the analyses reported in the article "independently" without the need for additional   information.

Note: The principle of reproducibility applies here. In other words, rerunning the same analysis on the same data should result in the same results. A good test of whether the documentation is sufficiently specific for peers is to have one of the co-authors rerun the analyses (or part of them) as a check.

Accepted or published manuscript or publication

The data package must include a short explanation of the problem definition, research design, conceptual framework, data collection (sampling, selection, and representativeness of informants), and applied methods. An electronic version of the published manuscript is sufficient for this, but this can also be a unique referral, for example, a DOI.

Access and verification

The data package must be available on request to the Science Committee for evaluation. In principle, It must be available to all the co-authors of the article and, in any event, to at least two persons. The second person may be a co-author at or outside TSB. In the case of PhD research, this role may be performed by supervisors/co-supervisors, and, in the case of research by senior researchers, the manager may perform this role. Any restrictions on the use of research data, e.g., in follow-up research for persons with access may be laid down contractually but should not in principle present an obstacle to the verification of results in articles.

Note: The principle here is that co-authors share responsibility for the handling and statistical processing of the data and the reporting of the results. Ideally, in addition to the person with prime responsibility for analysis and the reporting of results, the co-author should replicate the analyses and each author should verify the other one's work using the 'co-pilot' system.

Retention period

The data package should be retained and remain available for at least ten years following the definite publication date of the article. As far as possible, anonymized research data are made public subsequent to completion of the research in alignment with the [FAIR principles](#).

Data storage

The data package must be stored digitally in a secure and robust manner with automatic backups as the default.

Note: 'Secure and robust' means that data cannot be stored on a single storage device (e.g., a single hard disk, DVD, or USB stick). Researchers are expected to opt for the most robust and secure option available for hosting the data package (e.g., surfdrive, DataverseNL).

Checklist data package

Metadata and data collection
- Logbook including the date each file was created, who created that file, and who edited the file and when it was edited
- Who had/has access to the package
- Ethical review
- Role of each author/co-author
- Who collected the data and where and when
- If an existing database is used: add original version and date
- In case of external finance/grants, specify who provided these
- Date of acceptance manuscript (include reference, e.g. DOI)

Raw Database
- All anonymized data collected during the study. Not included: information that identifies persons and/or consent forms
- In case of non-digital materials, specify where this is stored and how the data can be traced.
- In case the data is stored elsewhere in an external archive, add the source reference.

Material
- All digital material is included in the data package

Statistical processing
- Syntax / Scripts / Code (e.g. Python or R)

Processed database
- Sufficiently documented and processed database

Accepted or published manuscript or publication
- Problem definition
- Research design
- Conceptual framework
- Data collection
- Methods applied

→ The data package must be available on request to the Science Committee for evaluation.
→ The data package should be retained and remain available for at least ten years.
→ The data package must be stored digitally in a secure and robust manner with automatic backups as default.