

**Guideline on Data Handling and Methods Reporting (DHMR)**

TSB (School of Social and Behavioral Sciences) Science Committee,  
2017

The origin of the Guideline lies in the principle of 'verifiability' laid down by the Association of Universities in the Netherlands (VSNU, 2014) as shown in the box below. The recommendations of the Levelt and Schuyt committees (Levelt Committee, Noort Committee and Drenth Committee, 2012; KNAW, 2012) and best practice have also been taken into account.

**Box 1: The VSNU principle of verifiability***III. Verifiability**Principle*

*Presented information is verifiable. Whenever research results are published, it is made clear what the data and conclusions are based on, from where they originate and how they can be verified.*

*Definition*

*Conduct is verifiable when it is possible for others to assess whether it complies with relevant standards (for instance of quality or reliability).*

*Elaboration*

*III.1. Research must be replicable in order to verify its accuracy. The choice of research question, the research set-up, the choice of method and the references to sources used are accurately documented in a form that allows for verification of all steps in the research process.*

*III.2. The quality of data collection, data input, data storage and data processing is closely guarded. All steps taken must be properly reported and their execution must be properly monitored (lab journals, progress reports, documentation of arrangements and decisions, etc.).*

*III.3. Raw research data are stored for at least ten years. These data are made available to other academic practitioners upon request, unless legal provisions dictate otherwise.*

*III.4. Raw research data are archived in such a way that they can be consulted at all times and with a minimum expense of time and effort.*

*III.5. The source of all educational material, written as well as oral, is stated.*

Source: VSNU, 2014, pp. 8.

## Implementation

TSB staff and PhD students must ensure that there is a data package for each empirical article or book/chapter of which they are author and each empirical chapter of a thesis on which they are taking their PhD (even if the TSB employee has not himself analyzed the research data).

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.

The data package must comply with eight requirements.

- (1) **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. The metadata must also indicate whether an ethical review has been carried out and give details (the protocol number, instance) of the review. The role of each author/co-author must also be described. In addition, the data package must include detailed information on who collected the data, where and at what location. 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 origin, version and date must be specified. In case of external finance / grants, it must be specified who provided these.

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 a digital logbook will therefore need to be created in some cases. 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.

- (2) **Raw database:** The data package must contain the 'raw database' as initially available to the TSB staff member digitally or in digitizable form (e.g. scans of paper questionnaires). This must be a read-only file that is time marked and cannot be edited after storage. The raw database should be fixed upon

completion of the data collection. The date of 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 or consent forms (which must also be retained unless other arrangements have been made on account of ethical objections). In the case of non-digital source material (physiological findings, audio recordings, paper questionnaires, etc.) it must be stated where this is stored and how the data can be traced. In cases where the TSB researcher is not allowed to have 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 the case of such data management limitations (e.g. by means of a mention in the author's note to the article).

Note: The raw database is the first database obtained digitally by the TSB staff member. Ideally this should be first digital file created by anyone, but in cases where datasets/secondary datasets from third parties or files e.g. downloaded from data repositories or other types of publicly available databases are used this file may also mean these. Self-collected data means data such as entered 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 reasoned, with documentation, in a file in the data package. In such cases the aim should be to include in the data package a random sample from the original database or a screenshot of part of the raw data.

- (3) **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. one hard disk, DVD or USB stick). Researchers are expected to opt for the most robust and secure option available for the hosting of the data package.

- (4) **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. The research must in principle be able to be replicated by a peer, based on the documents in the data package without the need for additional information.

- (5) **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 at (6).

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.

- (6) **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: Again, the principle of replicability applies here. A good test of whether documentation is sufficiently specific for peers is to have one of the co-authors replicate the analyses (or part of them) as a check.

- (7) **Access and verification:** The data package must be available on request to the Science Committee for evaluation. It must in principle 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, there should be a co-author who

replicates the analyses and they should verify each other's work using the 'co-pilot' system.

- (8) **Retention period:** The data package should be retained and remain available for at least ten years following the definite publication date of the article.

Note: The VSNU guideline refers to ten years without specifying when this period commences. The American Psychological Association specifies the publication date of an article, thus making for a clear period.

### References

VSNU (Association of Universities in the Netherlands) (2014). *De Nederlandse Gedragscode Wetenschapsbeoefening* [The Dutch Code of Conduct on Scientific Practice]. The Hague, VSNU.

Commissie Levelt, Commissie Noort and Commissie Drenth (2012). *Falende wetenschap: De frauduleuze onderzoekspraktijken van sociaal-psycholoog Diederik Stapel* [Failing science: The fraudulent research practices of social psychologist Diederik Stapel]. Tilburg University.

KNAW (Royal Netherlands Academy of Arts and Sciences) (2012). *Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens* [Handling scientific research data carefully and with integrity]. Amsterdam, KNAW.