2022

MARCH

Guideline for the archiving of academic research for

Faculties of Behavioural and Social Sciences

in the Netherlands

Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences in the Netherlands

March 2022

Index

Preamble

The principles of honesty, scrupulousness, transparency, independence, and responsibility form the basis of research integrity (UNL, 2018). Abiding by these principles enlarges trust and quality of academic research, thereby improving its relevance to society. The current guideline is developed with input from all DSW faculties and offers guidance for the archiving of academic research published by researchers at the Dutch faculties of social and behavioural sciences, drawn from the principles of scrupulousness, transparency, and responsibility. The guideline seeks to improve archiving of social and behavioural research using both quantitative and qualitative methods, in order to safeguard continued availability of qualitative or quantitative research data, detailed descriptions of research materials and approaches, and an overview of the data processing and publication processes after the research has been published.

This guideline is not meant to replace other existing guidelines or regulations related to data management, open science, data processing agreements and privacy aspects in the design stage of a research project. The document can be seen as an initiative that is part of a broader effort to promote research integrity among researchers focusing on both quantitative and qualitative studies at faculties of behavioural and social sciences in the Netherlands. Rather than functioning as a strict straightjacket, it intends to provide a clear guideline, which can be further fleshed out under the motto 'apply or explain', taking into account existing regulations at the faculty or university level.

Researchers working in the social and behavioural sciences at a Dutch university will be held to these standards to ensure that research integrity in general and transparency in particular can be ensured. Given the various distinct methodologies of scholarly research carried out under the general "social science" header, there are two main approaches that can be identified and should be implemented to ensure scientific integrity and its future assessment. The first is primarily for quantitative research designs and quantitative data that can most often relatively easily be de-identified (pseudonymized or anonymized) and stored in a repository in full. The second is for scientific research that is structured by qualitative and interpretive research designs and epistemologies that generate data and information that may have a different character and most often cannot be de-identified and stored in an identical manner as quantitative data. Regardless of methodological approach, all researchers have an obligation to follow the standards of integrity and transparency set in this document. All researchers must be aware of the specific regulations that govern their type of research and adhere to these regulations[1] (except where motivated exceptions are allowed).

[1] For specific regulation regarding the ethical, legal and social implications of health-related research, researchers can consult the ELSI Servicedesk

1.1   Purpose of these guidelines

These guidelines for the archiving of academic research set out the preconditions for the archiving of data, materials and information that form the basis for publications – in other words, (descriptions of) data, materials and information that are needed in order for academic peers and other consumers of the research to replicate, reproduce, and/ or assess the published research results. These guidelines relate to the data, materials and information with respect to publications that appear in their definitive form as of 1 September 2021.[2] The guidelines are based on the principle of retroactive accountability, i.e. reporting after a publication has appeared. The norm behind these guidelines is that each researcher is responsible for archiving data, materials and information, and the publications based on them, in a responsible and transparent way, in order to keep the data for future verification or checking by academic peers, and re-use. In situations where this document does not provide clear-cut rules, researchers are expected to act in the spirit of these guidelines rather than observing them to the letter.
Faculties will be expected to apply these national guidelines. The guidelines will be evaluated every two years, under the responsibility of the deans of the faculties of social and behavioural sciences (DSW).

1.2   To whom do these guidelines apply?

These guidelines apply to all faculty staff members who conduct research in the context of a temporary or permanent employment contract, all PhD candidates who conduct research under the supervision of a professor, and all research master's students. The guidelines do not apply to bachelor's and one-year master's students, unless their research results in an academic publication. Research conducted by bachelor's and one-year master's students falls under the formal responsibility of their supervisors.

All researchers at the faculty must adhere to *The Netherlands Code of Conduct for Research Integrity*.[3] These guidelines are a concrete embodiment of the principle of transparency and the related norms set out in the UNL Code of Conduct. The Netherlands Code of Conduct also requires researchers to make data as open as possible after publication or to document valid reasons for not sharing the data.

---

[2] Originally, around 2017 and 2018, this document was the result of the efforts of a committee established to this end by the DSW, consisting of Marc van Veldhoven (UvT, later replaced by Jelte Wicherts), Rob Eisinga (RU), Rosanne Janssen (UM) and Peter van der Heijden (UU). This latest version has been edited by the DSW committee Scientific Integrity, data storage and reproducibility, consisting of Peter van der Heijden (UU), Sander Nieuwenhuis (UL), Jelte Wicherts (UvT) and Esther Hoorn (RUG), using suggestions of a group of qualitative researchers of the UL (Wolfgang Kaltenbrunner, Marianne Maeckelbergh, Joop van Holsteijn and others).

[3] Netherlands Code of Conduct for Research Integrity, Standards for good research practices. https://doi.org/10.17026/dans-2cj-nvwu

1.3   Raw data, personal data and research data

Within the framework of the transparency and replicability of research, raw data must of course be retained. Raw data are the unedited data that are collected within the framework of a research project, for example:

- Registrations derived from experimental research
- Survey data from questionnaires completed within the framework of research (including longitudinal research), collected by the researcher themselves or by an external fieldwork organization
- (Transcripts of) video material collected within the framework of qualitative research (open interviews, observations)
- Notes taken within the framework of qualitative research or research using source material

Raw data must always be de-identified as soon as and insofar possible so that they cannot be directly traced back to people or groups of people. Data that can be directly or indirectly traced back to a person are known as personal data. This includes not only name and address details, but also photographs, audio - and video material, and other identifying information. The de-identified raw data and the personal data together form the research data.

Guidelines concerning publication packages

These guidelines relate to all research publications listed in the faculty's academic annual report. In order to ensure the transparency of qualitative and quantitative empirical research, all information that is needed to be able to assess the results must be archived (in English). This information is stored in a 'publication package'.

2.1 What must be stored in a publication package?

We make a distinction between publication packages resulting from quantitative research and from qualitative research projects, while noting the existence of mixed methods that employ both qualitative and quantitative elements and should be handled according to their main focus.

2.1.1 Quantitative research

The following materials must be stored for each published empirical study (article, volume, book chapter, PhD thesis chapter, Research Master's thesis, consultable internal report, etc.):

1. The published (or accepted) manuscript or publication.
2. A brief description of the problem definition, research design, data collection (sampling, selection and representativeness of informants) and methods used. An electronic version of the published manuscript will generally suffice.
3. The instructions, procedures, the design of the experiment and stimulus materials (interview guide, questionnaires, surveys, tests) that can reasonably be deemed necessary in order to replicate the research. The materials must be available in the language in which the research was conducted. The publication package must be in English.
4. When using primary data, the (de-identified) raw data files (providing the most direct registration of the behaviour or reactions of test subjects/respondents, for example an unfiltered export file of an online survey or raw time series for an EEG measurement, e-dat files for an E-Prime behaviour experiment, recordings or transcripts of interviews, descriptions of observations, archive and other source or media material). Documentation of the steps taken to de-identify the data and a blank consent form. If the raw data files have been accessibly stored in an external archive (such as storage facilities at DANS), making reference to the files in this archive will suffice. Such externally archived raw data may include primary or secondary data. Raw data may not be changed once they have been made digitally available.
5. Computer code (for example Atlas.ti, SPSS/JASP syntax file, MATLAB analysis scripts, R code) describing the steps taken to process the raw data into analysis data, including

brief explanations of the steps in English, for example a brief description of the steps taken in the qualitative analysis of primary research data, i.e. themes, domains, taxonomies, components.

6. The data files (either raw or processed) that were eventually analysed when preparing the article (e.g. an SPSS data file after transforming variables, after applying selections, etc.) The latter is not necessary if the raw data file was directly analysed.

7. Computer code (for example syntax files from SPSS/JASP, Atlas.ti, Matlab, R; syntaxes of tailored software) describing the steps taken to process the analysis data into results in the manuscript, including brief explanations of the steps in English.

8. The data management plan.

9. A readme file (metadata) describing which documents and files can be found where and how they should be interpreted. The readme file must also contain the following information:

   a. Name of the person who stored the documents or files
   b. Division of roles among authors, indicating at least who analysed the data
   c. Date on which the manuscript was accepted, including reference
   d. Date/period of data collection
   e. Names of people who collected the data
   f. If relevant: addresses of field locations where data were collected and contact persons (if any)
   g. Whether or not an ethical assessment took place before the research, and, if relevant, study reference from and statements made by the Ethics Review Committee
   h. Whether the data is made open or not and if not, a valid reason for not opening up the data

   The readme file must be sufficiently clear. A relevant fellow researcher must be able to replicate the results discussed in the publication based on the components of the publication package.

10. Documents related to the ethical approval or a reference to such documents.

2.1.2 Qualitative research

For qualitative, interpretative methodologies, a distinction should be made between the two main criteria for research integrity, i.e., transparency and reproduction. Transparency is a valid and legitimate demand also for qualitative research (and data), but reproduction is not considered possible in all cases, due to the very nature of the research designs and epistemology. Qualitative data are often impossible to fully de-identify and the research data is often gathered in forms and formats that cannot be stored in a digital repository.

Of course, some of these data may be highly sensitive and cannot be shared with others without breaking ethical rules and the confidentiality that is often guaranteed to informants and other (human) sources of information. But as the aim of these guidelines is not sharing data but storing data, qualitative research should also be archived. Sensitive data should be stored on secured faculty servers. And when the format does not allow researchers to store original objects, it suffices to store pictures of the material. These data should be stored safely in a way that is accessible to the researcher who gathered the data.

Researchers are therefore expected to store their data safely and to make specific plans for the time period of storage of their data, where and in which manner the data will be stored, and what will be done with the data once the research project ends or, for long-term ongoing research, once the researcher retires from research reporting etc. This calls for an elaborate and transparent data management plan or another, similar or equivalent form of data storage plan that describes: what kind of data will be gathered, by whom, in what format, where and in which form these will be stored, and to what extent and under what conditions this data will be shared and with whom, and any specific steps that will be taken to share the data that is safe to be shared. The researcher should be aware that according to the Netherlands Code of Conduct for Research Integrity there may be (highly exceptional) cases in which there are compelling reasons for components of the research, including data, not to be disclosed to an investigation into alleged research misconduct. Such cases must be recorded and the consent of the board of the institution must be obtained prior to storing the components and/or data in question. This documented exception must also be mentioned in any results published.[4]

---

[4] Netherlands Code of Conduct for Research Integrity, Standards for good research practices, 3.2 Design, 12 B. https://doi.org/10.17026/dans-2cj-nvwu

In addition to safely storing data, the (qualitative) researcher shall make sure to maintain a record of the following metadata:

1. The dates that the researcher carried out the data collection (e.g. dates of interviews or observation, period(s) of time spent in the field (start date and return date), etc.;
2. The type of activities carried out (e.g., participant observation, number of interviews, frequency and character of observation, familiarizing oneself with the field, informal and formal conversations, other types of recording activities);
3. Interview and observation guides (if available);
4. Any hard evidence of the period of time spent in the field (e.g. flight reservations, train tickets, etc.).

Researchers should be aware that they should be able to easily retrieve the above information upon request.

2.2   When must a publication package be stored?

A publication package must be stored within one month after the definitive publication of the manuscript. A publication package must be stored for each submitted research master's thesis. A publication package must be stored for each empirical chapter of a PhD thesis submitted to the thesis committee (or one single publication package if the thesis is a monograph).

Once a publication package has been stored, it will be fixed and can then no longer be modified (read only).

2.3   Who is responsible for storing publication packages?

If the first author works at one of the faculties of behavioural and social sciences, they will always be responsible for the archiving of the publication package, i.e. the storage of raw and edited data, syntax and materials, and additional information about the publication process as discussed above. Second or later authors who work at a faculty of behavioural and social sciences must know that the data have been carefully stored and how this has been arranged. This is particularly relevant if the first author does not work at a faculty of behavioural and social sciences.

If the first author works at one of the faculties of behavioural and social sciences, the second or later author may assume that the first author will follow the guidelines of his or her own university, and the second or later author will not have to create a publication package.

For PhD candidates and research master's students, the primary supervisor or the day-to-day supervisor respectively are responsible for storing publication packages. The primary supervisor or day-to-day supervisor may delegate the execution of this task, but they will continue to bear final responsibility.

In collaborative projects a specific plan to clarify responsibilities related to the data after the project might be required. The person who coordinates the research programme that covers the publication (which, depending on the faculty in question, could be a professor, head of programme or head of department) is ultimately responsible.

Adherence to the guideline will be discussed in performance and appraisal interviews. Formal final responsibility lies with the dean.

2.4   Who has access to the publication package?

Publication packages should be accessible by more than one researcher. The first author will have reading rights, but no right to delete or change versions. The first author will have writing rights for adding new versions. If a faculty has appointed a 'co-pilot' to check the analysis or a data steward to consider data management compliance, they will also be assigned reading rights. The faculty board can assign reading rights to a specific official to prepare for audits of publication packages on its behalf, for example, the coordinator of a research programme or a member of an academic integrity committee. After publication, academic peers should be granted access to the publication package if they make a reasonable request to verify or examine the published research results in the context of academic debate.

Guidelines concerning the storage of research data and documentation

3.1 Minimum storage period

For the retention period regarding research, a distinction is made between research data (and software) and the documentation of the process that has been carried out.

Publication packages must be centrally stored on a secure faculty server facility for at least 10 years after the publication appeared. In the event of research (or secondary research) data including personal data, the principle of data minimization (conform GDPR regulation) must be applied as soon as possible. The Netherlands Code of Conduct for Research Integrity offers options to deviate from the retention period of 10 years. However, in that case the raw and processed data must be saved for a period suitable for the discipline and the methodology. The following could be taken into consideration when deciding on the retention period:

- the nature (and especially the privacy sensitivity) of the data;
- the need for source material to substantiate the results;
- the applied scientific value of the research results;
- the effort to make the data available for re-use;
- the efforts of long-term preservation;
- the usefulness of source material for follow-up research.

The retention period of data management plans and data management protocols of projects, faculties and research institutes is at least 10 years, but not shorter than the retention period of the dataset.[5] These documents primarily relate to policy making, execution and financing of research, and quality assessment. Also included here are the (legal) advice of ethical committees and evaluations and further agreements with research partners.

3.2 Data minimization and retention

Data that can be traced back to individuals may in principle not be linkable to research data when this is no longer necessary for the purposes of the study. These personal data must be destroyed once they are no longer necessary for the purpose for which they were collected. Some specific studies may require retention of data that can be traced back to individuals, for example for the purpose of follow-up research or for longitudinal studies. Technical and organizational measures to protect the rights of data subjects need to be

---

[5] https://www.nationaalarchief.nl/archiveren/kennisbank/selectielijst-universiteiten-en-universitair-medische-centra-2020 [accessed March 18, 2021].

documented and will preferably be standardized for specific research scenarios. Protecting the right of data subjects is particularly important for raw data that cannot be de-identified (for example, video- and audio data).

One complicating factor lies in the wish to retain personal data for the purpose of reviewing the integrity of the research itself, for example to check whether the participants did indeed participate in the research. If such integrity reviews are regarded as part of the research whose integrity is reviewed and considered necessary in the field it is allowed to store data that can be traced back to individuals for this purpose. When research is published, such personal data must be stored separately; not in the publication package. As an alternative option, researchers, faculties and research institutes can develop a protocol to monitor the integrity of the research before archiving, after which the personal data can be deleted. It is not necessary to store the personal data for the sole purpose of enabling participants to exercise their rights under the GDPR.

The head of the relevant department or research program is responsible for monitoring the destruction of the research data on the required date. Official final responsibility lies with the dean.

3.3 How are storage and archiving of research data arranged?

The raw de-identified data must be saved on a faculty server that satisfies the relevant requirements for data storage in terms of security, robustness and automatic back-up facilities. The recommendation is to save the raw data in read-only format, before the data are made available for processing. Raw data stored in this way become fixed, which means that researchers will no longer be able to modify them deliberately or by accident.

All data that can be traced back to individuals must be stored on a second faculty server, which is physically separate from the first faculty server and thus from the raw data. If a key is required to link pseudonymized raw data to the personal data, this key must be stored on the second faculty server. This includes raw data that cannot be de-identified and must be stored, such as audio- and video data in its original format that cannot be transcribed.

External storage of raw data, for example in national or international data archives such as DANS – which makes the data publicly available, retrievable and citable – is recommended and in some cases required, for example when NWO requires this in a contract. However, this does not relieve researchers of their duty to store the data internally on the first faculty server.

Individual storage on an own hard drive, USB stick or cloud solution such as Dropbox does not suffice. Data that are collected within the framework of PhD [6] or postdoc research must be archived in such a way that continuity is ensured when the PhD candidate or postdoc in question leaves the faculty.

These storage requirements do not apply to sections of raw data that are managed by external organizations. Researchers who use data from external organizations must verify that the organization in question stores its data in accordance with a protocol that satisfies the requirements of these faculty guidelines.

---

[6] Each individual section of a PhD thesis (or the thesis as a whole) officially counts as a publication, even if it has not been published as such in a journal.
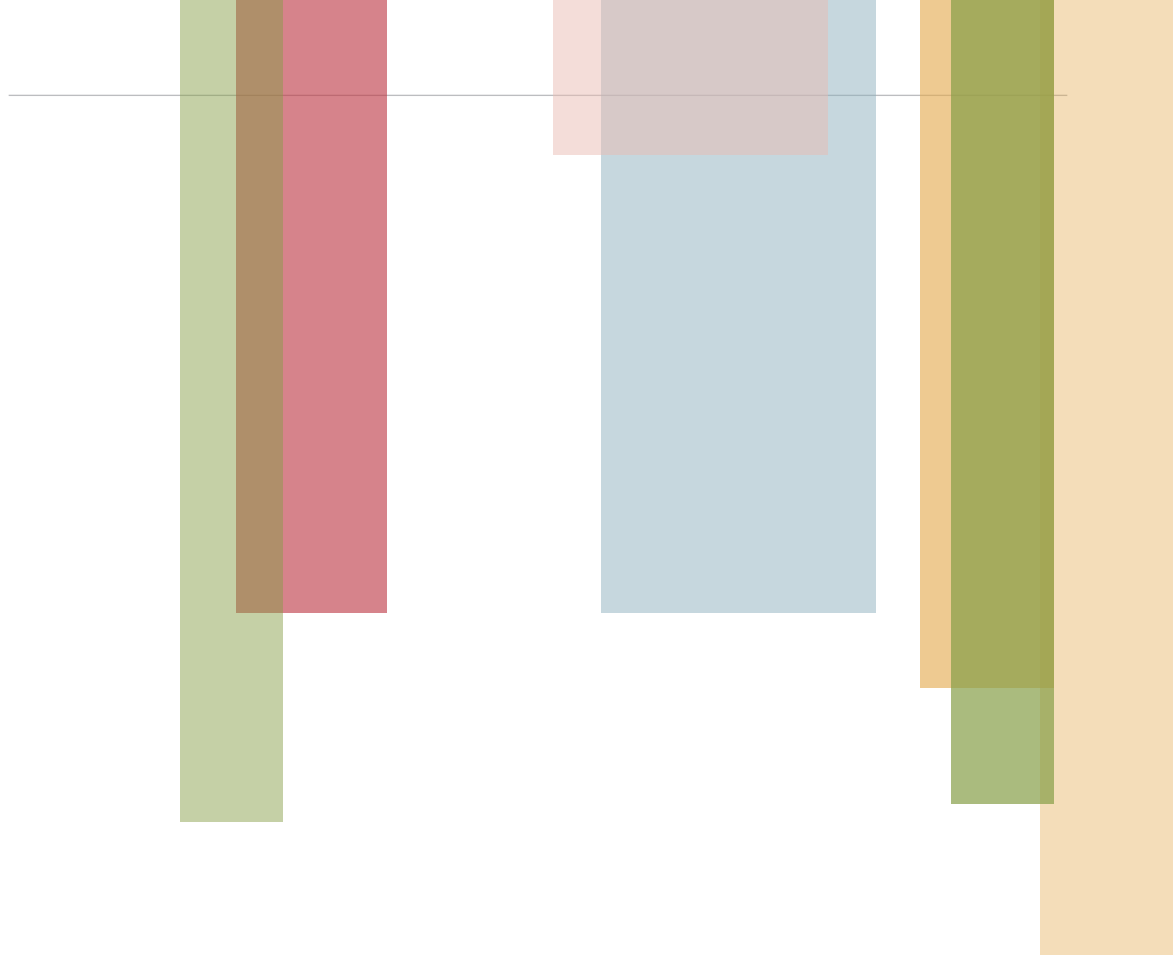
Faculty-specific policy

Individual faculties can choose to add the following rules to the above-mentioned guidelines concerning publication packages and storage of raw data:

1. Faculties may decide that the guidelines also apply to data collected within the framework of one-year master's and bachelor's research projects. The supervisor can then be appointed as the responsible party.
2. Faculties may decide to extend these guidelines to include storage of all data, including research that has not been published. This must be set out in a data management plan.
3. Faculties may define rules concerning ownership of data, for example that storage of data in a publication package will not result in a change of ownership.
4. Faculties may decide to make random inspections to check the existence and quality of publication packages.
5. Faculties may use different time periods and, for example, indicate that a publication package must be archived upon acceptance (rather than publication) of a manuscript.
6. Faculties may decide that each manuscript must state where the data are stored (a data statement) and which roles the various authors played.

Colophon