

Guidelines for the archiving of academic research for faculties of Behavioural and Social Sciences in the Netherlands Version 2.1, April 2018

1. Preamble

Discussions on academic integrity mainly focus on the transparency and replicability of research. Both these factors require access to qualitative or quantitative research data, detailed descriptions of research materials and approaches, and an overview of the data processing and publication processes. This document provides guidelines for archiving academic publications as well as the information needed to replicate the results discussed in such publications. This document thus relates to the archiving of published academic research and should not be regarded as guidelines concerning data management, data processing agreements and privacy aspects. The document can be seen as an initiative that is part of a broader effort to promote academic integrity among researchers focusing on quantitative and qualitative studies at faculties of Behavioural and Social Sciences in the Netherlands. Rather than functioning as a strict straitjacket, it intends to provide a clear orientation, which can be further fleshed out under the motto 'apply or explain' for each individual faculty depending on its circumstances.

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 to replicate research results, as well as their storage. These guidelines relate to the data, materials and information with respect to publications that appear in their definitive form as of 1 June 2018.ⁱ The guidelines are based on the principle of retroactive accountability, i.e. reporting after a publication has appeared. The principle 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 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 the national guidelines. Phased introduction of the guidelines will, however, be permitted in order to give researchers some time to familiarize themselves with the preconditions.

1.2 Existing frameworks within which these guidelines will function

- Personal Data Protection Act and, as of 25 May 2018, the General Data Protection Regulation
- Copyright Act, Patents Act, Databases (Legal Protection) Act
- Medical-ethical research protocol requirements derived from the Medical Research (Human Subjects) Act (WMO: *Wet Medisch-wetenschappelijk Onderzoek met mensen*), reviewed by a Medical Ethical Committee
- Collective Labour Agreement for Dutch Universities (CAO-NU)
- VSNU (Association of Universities in the Netherlands) Code of Conduct
- Local university policy frameworks

1.3 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 Scientific Practice*.ⁱⁱ These guidelines are a concrete embodiment of the 'verifiability' aspect set out in the *VSNU Code of Conduct*.

2. 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 and replicability of qualitative and quantitative empirical research, all information that is needed to be able to replicate 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?

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, conceptual framework, 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 (topic list, interview guide, questionnaires) 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 anonymized 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). 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 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, 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. 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, statements made by the Ethics Review Committee.
9. 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. If available, documents relating to the ethical approval.

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, he or she 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 he or she will continue to bear final responsibility.

2.4 Who has access to the publication package?

The first author will have reading rights. If a faculty has appointed a 'co-pilot' to check the analysis, he or she 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.

2.5 Storage period and central storage of publication packages

Publication packages must be centrally stored in a secure Faculty server facility for at least 10 years after the publication appeared. The reading rights are defined in Section 2.4.

3. Guidelines concerning the storage of raw 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 him or herself 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 should preferably be stored digitally. Stored raw data must always be anonymized so that they cannot be directly traced back to people or groups of people.^{iii,iv} Data that can be directly traced back to a person are known as personal data. This includes not only name and address details, but also photographs and video material. The raw data and the personal data together form the research data.

3.1 Minimum storage period for research data

The Faculty follows the guidelines set out by the VSNU and APA with regard to the storage period for raw data: 'minimum of 10 years after the publication of this research'. No maximum storage period applies to stored raw data that has been anonymized; these data may be kept longer than necessary for the purpose for which they were collected or processed (in accordance with Section 10.2 of the Personal Data Protection Act).^v

One exception to this rule applies to the storage of personal data for medical files, and therefore also data derived from research that falls under the Medical Research (Human Subjects) Act (WMO). These data must be stored for at least 15 years, in accordance with Article 454.3 of the Medical Treatment Contracts Act (WGBO: *Wet op de geneeskundige behandelingsovereenkomst*).

3.2 Maximum storage period for research data

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 the purpose for which they were collected has been achieved, in accordance with Section 10.1 of the Personal Data Protection Act. 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. In such cases, these personal data may be stored in accordance with Section 10.2 of the Personal Data Protection Act.

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

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, the Act appears to allow the storage of 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.

3.3 How are storage and archiving of research data arranged?

The anonymized raw 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.

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 the anonymized raw data to the personal data, this key must be stored on the second Faculty server.

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^{vi} 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.

3.4 Who is responsible for storing research data?

The researcher him or herself is primarily responsible for adhering to these guidelines concerning the storage of raw data. When it comes to PhD candidates and Research Master's students, the primary supervisor or day-to-day supervisor respectively are responsible. 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 guidelines will be discussed in performance and appraisal interviews. Formal final responsibility lies with the Dean.

4. 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:

- (i) 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.
- (ii) 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.
- (iii) 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.
- (iv) Faculties may decide to make random inspections to check the existence and quality of publication packages.
- (v) 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.
- (vi) Faculties may opt for a phased introduction of the national guidelines.
- (vii) Faculties may decide that each manuscript must state where the data are stored (a data statement) and which roles the various authors played.

ⁱ This document is 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).

ⁱⁱhttp://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf

ⁱⁱⁱ Personal data means: any information relating to an identified or identifiable natural person (*Personal Data Protection Act* Section 1.a).

^{iv} *Personal Data Protection Act*, Section 10

1. Personal data may not be kept in a form which permits identification of a data subject for longer than is necessary for the purposes for which the data were collected or for which they are further processed.

2. Personal data may be kept for longer than provided in subsection 1 in so far as they are kept for historical, statistical or scientific purposes and the controller has taken the measures necessary to ensure that the data concerned are used solely for those specific purposes.

Section 11

1. Personal data may be processed only in so far as they are adequate, relevant and not excessive in relation to the purposes for which they are collected or further processed.

2. The controller will take the measures necessary to ensure that personal data are correct and accurate in relation to the purposes for which they are collected or further processed.

^v However, Article 30, 'Academic research and statistics', of the *Vrijstellingsbesluit Wet Bescherming persoonsgegevens* [Exemption decree concerning the Personal Data Protection Act] applies:

-
1. Section 27 of the Act does not apply to data processing by institutions for academic research or statistics that exclusively serve research conducted or to be conducted by them, to the extent that this data processing satisfies the requirements set out in this article.
 2. Data are only collected, processed and checked for the purposes of a specific study or specific statistics.
 3. The personal data to be processed will only include:
 - a. surname, given names, initials, titles, sex, date of birth, address, postcode, place of residence, telephone number and similar details that are needed for communication purposes, as well as the bank account number of the person involved
 - b. an administration number that bears no information
 - c. data other than those referred to under a and b, which are acquired for the purposes of a specific study or specific statistics.
 4. The personal data will only be supplied to:
 - a. parties, including third parties, who are responsible for the activities referred to in Article 30.2 or for the management thereof, or who are necessarily involved in such activities
 - b. other parties, for those cases referred to in Section 8 a, c and d and Section 9.3 of the Act.
 5. The personal data referred to in Article 30.3 a, with the exception of sex, place of residence and year of birth, must be deleted no later than six months after the acquisition of the information about the person involved as referred to in Article 30.3 c.

^{vi} 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.