

FACULTY OF SOCIAL AND BEHAVIOURAL SCIENCES RESEARCH DATA STORAGE (ARCHIVING) PROTOCOL 2016

REVISED VERSION OF *PROTOCOL ON RESEARCH DATA (EFFECTIVE PER SEPTEMBER 1st, 2014)*

(Further faculty details supplementing the Utrecht University Policy Framework for Research Data)

PURPOSE OF THIS PROTOCOL

This research data protocol for 2016 lays down the conditions within the Faculty of Social and Behavioural Sciences for storing (archiving) research data. The protocol concerns publications that appear starting January 2016.ⁱ The basic principle of this protocol is that every researcher is personally responsible for ensuring that his or her research data is stored responsibly and transparently. If this protocol lacks specific guidelines for certain cases, researchers are expected to act in the spirit of this protocol.

Divisions (departementen) and Departments (afdelingen) are at liberty to formulate more detailed specifications within these faculty-wide conditions. If a given department feels the need to significantly depart from this protocol, it must submit a substantiated request to this effect to the Faculty Board, which will then seek advice from the Committee for Academic Integrity (*commissie wetenschappelijke integriteit*).

NEED FOR THIS PROTOCOL

Transparency and reproducibility play an important role in discussions relating to academic integrity. Both of these assume access to data – under certain conditions to be agreed with the author concerned – and the ability to reproduce the research results on the basis of existing data.

EXISTING FRAMEWORKS WITHIN WHICH THIS PROTOCOL OPERATES:

- the Personal Data Protection Act (*Wet bescherming persoonsgegevens*);
- the Copyright Act (*Auteurswet*) and the Netherlands Patent Act (*Rijksoctrooiwet*);
- the medical and ethical research protocol requirements set forth in the Medical Research Involving Human Subjects Act (*Wet Medisch-wetenschappelijk onderzoek met mensen*, WMO) and assessed by the Dutch Medical Ethics Assessment Committee (*Medisch Ethische Toetsingscommissie*, METC);
- the Collective Labour Agreement for Dutch Universities (CAO NU);
- the Code of Conduct of the Association of Universities in the Netherlands (*Vereniging van Nederlandse Universiteiten*, VSNU);
- the Utrecht University Policy Framework for Research Data (*Universitair Beleidskader Onderzoeksdata Universiteit Utrecht*);
- Information security policy Utrecht University.

TO WHOM DOES THIS PROTOCOL APPLY?

This protocol applies to everyone within the Faculty of Social and Behavioural Sciences who actively conducts research as part of a temporary or permanent employment contract, including all PhD candidates carrying out doctoral research under the supervision of professors and all Research Master's students. It does not apply to research conducted by Bachelor's and one-year Master's students. For the time being, research conducted by these students is formally the responsibility of their supervisors. The protocol applies to all publications that will be listed to in the 2016 Annual Report.

All of the faculty's researchers subscribe to the Netherlands Code of Conduct for Scientific Practice. The Education and Supervision Agreement signed by each PhD candidate states that he/she has read and subscribes to this code of conduct.ⁱⁱ

DATA TO WHICH THIS PROTOCOL APPLIES:

1. data collected and archived by external organisations (e.g. Statistics Netherlands (CBS), Netherlands Kinship Panel Study, European Social Survey, European Values Study, etc.);
2. survey and other data in the context of research (longitudinal or otherwise) collected by the researchers themselves or as part of work commissioned by an external organisation, such as TNS NIPO;
3. data collected for experimental research;
4. data collected for qualitative research (open interviews, observations, etc.).

N.B. It also relates to data that researchers have collected while working at other institutions.

DEFINITION OF RAW DATA

Raw data includes:

1. completed questionnaires and video materials, for example, which should preferably be stored digitally;
2. other data collection materials (e.g. stimulus materials and software used for data collection). These materials should be stored unless they are freely available elsewhere in an identical form. In that case, a description of the materials and a reference to the source will suffice.

MINIMUM DURATION OF DATA STORAGE

Any raw data stored are always anonymised and cannot therefore be traced directly to the individual concerned.^{iii,iv} With regard to the duration of raw data storage, the faculty follows the VSNU and APA guidelines: 'a minimum of ten years after the publication of this research'. An exception applies to the storage of personal data for a medical file and therefore also data originating from research subject to the WMO. This data must be stored for at least 15 years (in accordance with Article 454, paragraph 3 of the Medical Treatment Agreement Act (*Wet op de geneeskundige behandelingsovereenkomst*, WGBO)).

MAXIMUM DURATION OF DATA STORAGE

There is no maximum storage duration for raw data (which therefore cannot be traced to the individual concerned), which may be stored for longer than is necessary for the purpose for which they are collected and processed (pursuant to WBP Article 10, paragraph 2).^v By way of departure from this, data that *can* be traced to the individual may *not* be linkable to research data if this is no longer necessary for the purposes of the research and must be destroyed if those purposes have been achieved (pursuant to WBP Article 10, paragraph 1). An example of a case of data that are traceable to an individual being necessary for the purposes of the research is when conducting follow-up research or longitudinal studies, in which case the traceable data may continue to be stored (pursuant to WBP Article 10, paragraph 2). However, it is not permitted to store traceable data purely in order to verify whether participants actually participated in the research.

The head of the department concerned is responsible for ensuring that research data are destroyed on the appropriate date.

HOW IS THE STORAGE (ARCHIVING) OF RESEARCH DATA ORGANISED?

The anonymised raw data are stored on the faculty server, which meets the requirements to which data storage may be subject in terms of security, robustness (integrity and quality) and automatic backing up. Data that can be traced to the individual concerned are stored on the O-drive, separately from the raw data. If a key is needed to link the anonymised data to the individuals concerned, the key will be stored on the O-drive. An upload portal will be developed for this purpose.

External storage of raw data in national or international data archives, such as Data Archiving and Networked Services (DANS) – where the data is publicly available, findable and citable – is recommended and in certain cases compulsory, e.g. if the Netherlands Organisation for Scientific Research (*Nederlands Organisatie voor Wetenschappelijk Onderzoek*, NWO) demands this contractually, but this does not relieve the researcher of responsibility for data stored on the faculty server.

Individual storage on a private hard drive or USB, or 'in the cloud' (e.g. Dropbox) is not sufficient. Data collected by employees should always be archived in order to ensure continuity when employees leaves the faculty.

These storage requirements do not apply to (parts of the raw) data managed by external organisations or data involving secondary analysis. Any researcher making use of data from external organisations must ensure that this organisation stores its data in a way that corresponds with the requirements of this faculty protocol.

PRECONDITIONS FOR DATA PACKAGE (PREFERABLY IN ENGLISH) FOR EVERY EMPIRICAL ARTICLE, BOOK CHAPTER, ETC.

What needs to be stored?

For each published empirical study (article, chapter in a book, chapter in a thesis, Research Master's thesis, retrievable internal report, etc.):

1. very brief description of the hypothesis, research design, conceptual framework, method followed and structure of the data collected. As a rule, an electronic version of the accepted or published manuscript is sufficient for this;
2. instructions, procedures, the design of the experiment and any stimulus materials reasonably required for replication. The materials must be available in the language in which the research was conducted. An English translation is optional (but appreciated);
3. raw data files (the most direct record of the behaviour or responses of test subjects/respondents, e.g.: an unfiltered export of an online survey or raw time series for an EEG measurement, e-data files for an E-Prime behavioural experiment);
4. computer code (e.g. SPSS syntax file, MATLAB analysis scripts) that describes the steps to obtain the final analysis data from the raw data. This should include brief explanations of the steps in English;
5. data files that have been subjected to a final analysis for the purpose of writing the article (e.g. SPSS data file after transformation of variables, changes to selections, etc.). The latter is not necessary if the raw data file has been analysed;
6. computer code (e.g. SPSS syntax file) describing the steps followed to arrive at the results in the manuscript, based on the analysis data. This should include brief explanations of the steps in English;
7. a Readme file that describes where the documents and/or files can be found and how they should be interpreted. The Readme file should also contain the following information:
 - a. the name of the person who saved the documents and/or files;
 - b. the date on which the manuscript was accepted, including a reference;

- c. the date on/period in which the data was collected;
 - d. for each study, the names of the people who collected the data;
 - e. (if applicable) the addresses of any offsite locations where the data was collected and the relevant contact persons.
- The Readme file must be sufficiently clear to enable the results in the publication to be reproduced on the basis of the components of the data package.

WHO IS RESPONSIBLE FOR STORAGE?

If the first author is a researcher at the Faculty of Social and Behavioural Sciences, he/she is always responsible for the storage of raw and processed data, syntax, etc. and data that can be traced to the individual concerned, subject to the preconditions set forth in this protocol. If the second or later author is a researcher at the faculty, he/she will be informed that the data have been stored with due care and how they are stored. This applies in particular where the first author is not employed at by the faculty at Utrecht University. If the first author works at a Dutch university or at a university affiliated to the League of European Research Universities, the second or later author may assume that the first author follows the regulations of his/her own university and no data package needs to be stored by the second or later author.

For PhD candidates and Research Master's students, their supervisor or daily supervisor is responsible for data storage. The supervisor may delegate this responsibility for didactic reasons, but will still retain ultimate responsibility.

For each Research Master's student, a data package is stored on submission of the Master's thesis. The supervisor is responsible for storing data collected as part of a one-year Master's research or Bachelor's research.

For each PhD candidate, a data package is stored for every chapter of the manuscript (or a single data package for monographs) on submission to the Reading Committee.

WHEN

The data package is delivered shortly after final acceptance of the manuscript. The accepted version and the finalised publication are added later.

WHO IS ULTIMATELY RESPONSIBLE?

Primary responsibility for complying with this protocol lies with the researcher. For PhD candidates and Research Master's students, their supervisor or daily supervisor is responsible. Final responsibility rests with the holder of the chair under which a publication falls. Monitoring compliance with this protocol forms part of the Assessment & Development interviews. Formal responsibility lies with the dean of the faculty.

ⁱ The committee is also working on a protocol for storing research data during the data collection process.

ⁱⁱ http://www.vsnunl.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf

ⁱⁱⁱ Personal data refers to 'all information relating to an identified or identifiable natural person' (Article 1, paragraph a of the Personal Data Protection Act (*Wet Bescherming Persoonsgegevens*, WBP).

^{iv} WBP Article 10:

1. Personal data are no longer stored in a form that enables the individual concerned to be identified, unless this is necessary to achieve the purposes for which the data are collected or subsequently processed.
2. Personal data may be stored for longer than stipulated in the first paragraph insofar as this is for historical, statistical or scientific/academic purposes and the party responsible has made the necessary provisions to ensure that the data concerned is used exclusively for these specific purposes.

WBP Article 11:

1. Personal data should be processed only as much as is necessary for the purposes for which they are collected or subsequently processed, and is appropriate to the matter at hand, and never excessively.
2. The responsible party will take the necessary measures to ensure that personal data are correct and accurate, in view of the purposes for which they are collected or subsequently processed.

^v However, there is an Exemption Decree (*Vrijstellingsbesluit*) to WMP Article 30. Scientific research and statistics

1. WMP Article 27 does not apply to processing by organisations for scientific research or statistics exclusively for the purpose of research conducted or to be conducted by them, insofar as this processing meets the requirements stipulated in this article.
2. The processing takes place solely for collecting, processing and verifying data for a specific study or statistic.

-
3. No other personal data shall be processed except for:
 - a. surname, first names, initials, title, gender, date of birth, address, postal code, town or city of residence, telephone number etc. for the purposes of communicating the necessary data, as well as the bank and giro account number of the individual concerned;
 - b. information-free administration numbers;
 - c. data other than that referred to in a and b, acquired for the purposes of a specific study or statistic.
 4. The personal data are only issued to the following parties:
 - a. those persons, including third parties, charged with or managing the activities referred to in paragraph 2 or who are involved by necessity;
 - b. others, referred to in the cases in WMP Article 8, paragraphs a, c and d, and Article 9, paragraph 3.
 5. The personal data referred to in paragraph 3a, with the exception of gender, town or city of residence and date of birth, will be removed at the latest six months after the data referred to in 3c concerning the parties involved was obtained.