

Guidelines for the Archiving of Academic Research for Faculties of Behavioural and Social Sciences in the Netherlands March 2022

A Publication by the Committee Scientific Integrity, Data Storage and
Reproducibility from the Deans of Social Sciences in the Netherlands
(DSW)

With Local Guidelines for Implementation by VU Faculty of
Behavioural and Movement Sciences
2023-05-11; Version 3.4

Table of Contents

1. Preamble.....	2
1.1 Purpose of these guidelines	2
1.2 To whom do these guidelines apply?	3
1.3 Raw data, personal data and research data.....	4
2. Guidelines concerning publication packages	4
2.1 What must be stored in a publication package?	5
2.1.1 Quantitative research	5
2.1.2 Qualitative research.....	9
2.2 When must a publication package be stored?	10
2.3 Who is responsible for storing publication packages?	10
2.4 Who has access to the publication package?	11
3. Guidelines concerning the storage of research data and documentation:	11
3.1 Minimum storage period	11
3.2 Data minimization and retention.....	13
3.3 How are storage and archiving of research data arranged?	14
4. Faculty-specific policy	15

In the following sections, the guidelines published by the DSW at the national level are presented on the left, while the local implementation of these guidelines at the Faculty of Behavioural and Movement Sciences (FGB) at the Vrije Universiteit is presented on the right and is based on the principle “apply or explain”. FGB complies with the national guidelines, but where it does not comply, explanation for this deviation as well as further guidance is provided.

For question about these guidelines and the faculty’s implementation of them, please contact the FGB [Research and Policy Support Team \(REPS\)](#).

National Guidelines	Local Implementation (FGB)
<p>1. Preamble</p> <p>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.</p> <p>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.</p> <p>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 regulations¹ (except where motivated exceptions are allowed).</p>	<p>1.</p> <p>FGB complies with this introduction, with the additional caveat that the distinction between quantitative and qualitative data, and the manner in which these data should be handled, is not so clear cut. This is discussed further in the faculty’s implementation of these guidelines: in short, FGB argues that quantitative and qualitative data should not be handled differently. Differential handling of data has less to do with broad definitions of quantitative and qualitative and more to do with the specific kind of data being archived (e.g. administrative data, questionnaire responses, experimental measurements, (neuro)imaging, audiovisual recordings, textual data) and the privacy and/or security risks posed by the data, which are not necessary inherently greater just because data is qualitative. The faculty therefore sees these guidelines as a baseline standard for all research data regardless of whether it is considered qualitative or quantitative, and the faculty will provide additional support to researchers on its research support webpages for guidance on the specific handling of specific kinds of data.</p>
<p>1.1 Purpose of these guidelines</p> <p>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</p>	<p>1.1</p> <p>FGB complies with this requirement, however archiving may be appropriate on more occasions than simply after publication.</p>

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 2021ⁱⁱ. 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).

Data must always be archived after a research article is accepted for publication. Other situations where archiving may be appropriate include:

- Upon completion of a [research project](#), regardless of whether or not the data were used in a publication (especially when the data may be of use for a new project)
- Upon completion of raw data collection to ensure secure storage and to prevent loss or modification of the raw data

Archiving data in these cases is highly recommended, but not an absolute requirement. It is the responsibility of the research team to determine if the data should be archived in these situations.

These guidelines only address [data archiving](#), which is necessary to:

- Facilitate verification and replicability of research as required by the VSNU Code of Conduct for Research Integrity
- Meet legal requirements for medical research that is subject to the WMO law, the Good Clinical Practice (GCP) Guidelines and/or other medical research regulations.

[Publishing data](#) for reuse by third parties is not addressed in these guidelines.

This document is informed by the references cited at the end of the [FGB Research Data Management Policy](#), as well as the following regulations on the Good Clinical Practice (GCP) guidelines, and other medical research regulations, namely the Guidelines on Advanced Therapy Medicinal Products and the Medical Device Regulation:

- [Guideline on the content, management and archiving of the clinical trial master file \(paper and/or electronic\)](#)
- [Regulation \(EU\) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, repealing Directive 2001/20/EC](#)
- [Guidelines on Good Manufacturing Practice Specific to Advanced Therapy Medicinal Products](#)
- [Regulation \(EU\) 2017/745 of the European Parliament and the Council of 5 April 2017 on Medical Devices, Amending Directive 2001/83/EC, Regulation \(EC\) No 178/2002 and Regulation \(EC\) No 1223/2009 and Repealing Council Directives 90/385/EEC and 93/42/EEC](#)

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

1.2

FGB complies with this requirement

It is, however, recommended to have Bachelor's and one-year Master's students produce an informal data package to their research supervisors. This document does not need to be nearly as extensive as a formal data package, but should contain the datasets (raw and final), code/syntaxes used, any other

<p>master's students falls under the formal responsibility of their supervisors.</p> <p>All researchers at the faculty must adhere to <i>The Netherlands Code of Conduct for Research Integrity</i>ⁱⁱⁱ. 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.</p>	<p>documentation on data processing and analyses, and the final research paper. Research supervisors should determine the appropriate storage location and duration of storage for these data.</p> <p>NB: If Bachelor's or Master's students produce data that are used for a publication, then the data archiving guidelines below <i>do</i> apply. If these data are subject to the WMO law, the GCP guidelines or the other medical research regulations cited in section 1.1, supervisors must ensure that the data are stored for the duration required by these regulations.</p>
<p>1.3 Raw data, personal data and research data</p> <p>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:</p> <ul style="list-style-type: none"> • 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 	<p>1.3</p> <p>FGB includes additional detail on this requirement: in many cases raw data cannot be de-identified without irrevocably impacting the integrity of the raw data (such as with audiovisual data). The impact on the integrity of that data also depends on the extent to which the data will be de-identified. Raw data should only be de-identified to a level where the integrity of the original data can be maintained. Section 2.1.1 subsection 4 provides additional explanation on this.</p>
<p>2. Guidelines concerning publication packages</p> <p>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'.</p>	<p>2.</p> <p>FGB refers to this collection of information as a 'data package' rather than a 'publication package'. Earlier versions of these guidelines used the term 'archiving package' because archiving is often required in more instances than just after the publication of a research article and also to minimize confusion with the publication of the research data itself.</p> <p>FGB now refers to the collection of information as a 'data package' per 2023-05-11 to be in line with terminology used by archives and metadata schemas utilized at VU Amsterdam.</p>

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

FGB does not make a distinction in these guidelines between quantitative and qualitative data. The basic principles of archiving and what should be archived apply to both kinds of data. This will be discussed in more detail in [section 2.1.2](#)

The contents of a data package will differ depending on the purpose of archiving (see [section 1.1](#)). The following sections specifically describe what should be included in a data package after a research article has been published; for other purposes of archiving, some materials may not exist, such as a manuscript. This is not an issue for these other purposes so long as the data package contains whatever is necessary to correctly interpret its contents. Additionally, regardless of the purpose of archiving, a data package should provide enough information that one could repeat the process that produced the data. This does not necessarily mean that the data can be perfectly replicated, as this is extremely difficult in the social sciences whether the data are qualitative or quantitative. However, if the data package contains sufficient information on the data collection process, another researcher can replicate your workflow in order to produce similar data. Even for data that cannot be reproduced in any way, such as interview data, including the materials that explain the process of conducting and recording the interview contributes to transparency.

If data were initially archived upon the completion of data collection, and the data are subsequently used for a research publication, it is not necessary to archive the data again; one can simply refer back to the location where the raw data were initially archived. This is discussed further in [sections 2.1.1 subsection 4](#) and [3.3](#).

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.

2.1.1

Compliance or deviation at FGB is addressed for each requirement below. FGB also applies these requirements to both quantitative and qualitative data. Additional guidance on what needs to be archived for specific kinds of data, such as. survey data, audiovisual data, textual data etc. is being developed by the faculty. When this is ready, it will be found on the faculty's [Research Data Management support page](#).

1. This requirement depends on the reason for archiving. If there is no associated publication, a research protocol or research proposal should be archived.
2. This information does not need to be submitted separately in the data package if it is clearly described in the research publication, protocol or proposal.

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.

3. This information does not need to be submitted separately in the data package if it is sufficiently described in the research publication, protocol or proposal, excluding stimulus materials, which should be provided in full. Sufficiently described means that someone else could read the information and accurately carry out the same procedures

If several different languages were used in the stimulus materials, all versions used should be submitted in the data package. There should ideally be an English translation of all of these materials, even if an English version was not used to conduct the research (i.e. for purposes of greater transparency).

4. As stated in [section 1.3](#), raw data should only be de-identified if the integrity of the raw data can be maintained: in general, one can de-identify data up to step 4 of this [de-identification guide](#) without irrevocably impacting the integrity of the raw data; note that modifications to audiovisual data such as blurring or voice modification *is* an irrevocable change. It is also important to note that while this level of de-identification makes it more difficult to re-identify the raw data, this is not the same as anonymization. Lastly, once one begins [step 5](#) in the de-identification guide, data can no longer be considered raw data.

The identifiable data that has been separated from the raw data may need to be archived as well. See [sections 3.2](#) and [3.3](#) for the proper handling of identifiable data that has been separated from raw data.

NB: In some situations, the raw data to be archived are not in the possession of the FGB researcher who is creating the data package, e.g. research with business data that cannot leave the company's servers. If the researcher is unable to store the data in a VU archive, an agreement should be made with the owner of the data to ensure that the data will be stored appropriately for the agreed upon time frame, including an agreed upon location so that the data are findable upon request. The agreement must also allow for access to the archived data should verification or replication of the research results be necessary. If the data were also (pre)-processed by the external organization, the external organization must

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

either provide documentation of this processing so that it can be included in the data package submitted by the FGB researcher or access to this information must be included in the agreement that is made with the external organization.

The FGB researcher is still required to submit a data package with all of the other required components. The data package must also include the agreement made with the owner of the data so that someone reviewing the data package can determine where the data can be found and how they can be accessed.

5. FGB also requires that a non-proprietary copy of all code must be provided. Many programs have code/syntaxes that can be opened in a text editor (e.g. SPSS and SAS); if this is not possible a copy of the code must be provided in text format. In addition to the code, the program used and version number must be documented.

6. FGB also advises that any intermediate files created during the processing of raw data into processed data do not need to be archived, as long as the code showing these processing steps has been submitted (or, if this code has already been archived, a reference to the location of this code is included in the data package). If storage space is limited, the processed data do not need to be archived as long as the raw data, processing code and analysis code are included in the data package and the processed data can be regenerated with the processing code.

If the raw and/or processed data files are in a proprietary format (e.g. .sav, .xlsx, .doc), FGB strongly recommends that non-proprietary (e.g. .csv, .txt) copies of the data files are also archived, whenever possible.

7. See sub-section 5 above

8. FGB complies with this requirement.

9. FGB complies with this expectation. Researchers can download this [README.md template](#), which was designed to meet the national archiving requirements as well as any additional detail expected by the faculty. This README markdown file can be modified by using any text editor (e.g. Visual Studio

- 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 relating to the ethical approval or a reference to such documents.

Code, Atom, RStudio etc.) and the researcher can add as much additional detail as they feel is necessary.

NB: Some of the content in the README file *will* overlap with the metadata fields required when registering a project-level description of the archived dataset (see [section 4](#) for more information). Despite this overlap, a project-level description of the archived data must still be registered in order to make the archived data more [FAIR](#). The README file is also still necessary because it provides more detailed, narrative information. In the case of overlapping information, the same text can be used in both the project-level description and the README.

NB2: The README file does not need to be structured nor follow a specific metadata standard. The template included in these guidelines simply offers some structure to ensure that all of the required information is included, but the order and detail of the information can be modified by the researcher as required. Most importantly, the README file must be saved in an open format, such as markdown. If a researcher prefers not to use the markdown template provided in these guidelines, they should ensure that the README they create is saved in another open format such as .txt.

FGB agrees with this expectation. It is also recommended that FGB researchers use standard terminology as much as possible in the README so that readers can properly interpret the content of the data package.

10. FGB complies with this requirement

Addendum: Archiving Paper Documentation

NB: Currently, VU policy on the digitalization of paper documents prescribes that original paper documentation (e.g. informed consent forms, lab books etc.) must not be replaced by digital versions of these data sources. This means that even if paper documents are scanned and saved in a digital format, the paper documents *must not be destroyed* until the archiving term is complete. This policy may change if digitization techniques at VU Amsterdam are validated, at which point these guidelines will be updated.

NB2: For all studies subject to the WMO law, the GCP guidelines and/or the other medical research regulations cited in [section 1.1](#), original paper

documents **must not be destroyed** until the archiving term is complete, even if a digital copy is made.

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 longterm 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.^{iv}

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

2.1.2

FGB deviates from this requirement because the national guidance on the archival of qualitative data is insufficient. See [sections 2.1](#) and [2.1.1](#) above for more information on how qualitative data should be archived, i.e. with the same standard applied to quantitative data.

Additionally, FGB does not agree that reproducibility is an impossible goal with qualitative data.

Reproducibility is the process of reanalyzing the data already collected and confirming that the same results can be produced. If qualitative data are properly archived with sufficient documentation and metadata, reproduction is entirely possible. Replication (the process of replicating the results with new data) is more difficult, however another researcher could utilize the same procedures to carry out a new study and while they may not perfectly replicate the same results, they may be able to demonstrate consistent results, and potentially demonstrate generalization to other populations. And regardless of whether replication is possible, by archiving qualitative data in line with [section 2.1.1](#), researchers are transparent about the methods involved in the production of the data, which contributes to research integrity.

FGB further deviates from this section because both qualitative and quantitative data can be difficult to anonymize, however de-identification is actually feasible with qualitative data. It depends on how much identifying information can be removed from the data to still allow it to be useful for analysis. For more information see this [de-identification guide](#).

The secure storage of sensitive data also applies to both qualitative and quantitative data. Any sensitive, non-anonymous data needs to be stored in a secure archive. See [sections 3.2](#) and [3.3](#) for more information on the handling of identifiable data.

The requirement to make ongoing plans for the handling of the archived data during the archiving term also applies to both qualitative and quantitative data. See [sections 2.4](#) and [3.3](#) for more information.

<ol style="list-style-type: none"> 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.). 	
<p>2.2 When must a publication package be stored?</p> <p>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).</p> <p>Once a publication package has been stored, it will be fixed and can then no longer be modified (read only).</p>	<p>2.2</p> <p>FGB complies with this requirement with one exception: when archiving data used in a research publication, FGB researchers are expected to prepare a data package during the course their research and then submit the data package as soon as the research manuscript is accepted for publication by the journal.</p> <p>Additionally, with regards to other purposes for archiving:</p> <ul style="list-style-type: none"> • If data are to be archived upon completion of a research project, regardless of whether any publications were generated (see section 1.1), the data package should be submitted within one month of completion of the project. For particularly complex projects (spanning 10+ years and/or with 10+ different data sources), data packages may be submitted up to three months after completion of the project. • If an FGB researcher opts to archive the raw data upon completion of data collection (see section 1.1), the researcher can determine the timing of submission, although it is recommended to archive these data as soon as possible after data collection is complete.
<p>2.3 Who is responsible for storing publication packages?</p> <ul style="list-style-type: none"> ➤ 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 	<p>2.3</p> <ul style="list-style-type: none"> ➤ FGB complies with this requirement. Additionally, when archiving for purposes other than after a research article has been published (see section 1.1), the lead researcher for the project (e.g. project coordinator, primary investigator and/or project leader) is responsible for the data package. They may delegate the process to the researcher most directly responsible for the project, but the lead researcher remains ultimately responsible for ensuring the completeness and appropriate storage of the data package. ➤ FGB deviates slightly from this requirement: if the lead author is not an FGB staff member, the FGB co-authors should discuss and confirm with the lead author that they will archive the data, rather than assuming this. This applies regardless of

<p>or later author will not have to create a publication package.</p> <ul style="list-style-type: none"> ➤ For PhD candidates and research master’s students, the primary supervisor or the day-today 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. 	<p>whether the first author works at another faculty of behavioural and social sciences. Additionally, if the lead author is not an FGB staff member and does not have adequate data archiving facilities at their institution, the FGB co-authors should work together with the lead author to find an appropriate archiving solution.</p> <ul style="list-style-type: none"> ➤ FGB complies with this requirement. In addition, if FGB supervisors delegate this task to their candidate/student, they are expected to inform the candidate/student about these guidelines at the start of the PhD/Master’s project so that the candidate/student can appropriately plan for data archiving upon completion of the project. The supervisor is ultimately responsible for ensuring the completeness and appropriate storage of the data package. ➤ FGB complies with the requirement. Specific plans for collaborative projects should be drawn up to confirm and document the responsibilities for archiving. This is also discussed further in section 2.1.1 subsection 4 regarding data that is not in the possession of FGB researchers. ➤ FGB complies with this requirement
<p>2.4 Who has access to the publication package?</p> <p>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.</p>	<p>2.4</p> <p>FGB complies with this requirement, with some additional considerations.</p> <p>When data are archived for purposes other than in relation to a research publication (see section 1.1), the lead researcher for the project should have reading rights to the data package and they should ensure that at least one other person has reading rights.</p> <p>FGB will allow the lead author/lead researcher to determine which other individual should also have access to the data package.</p> <p>Should both individuals who have access to the data package no longer work at VU Amsterdam during the archiving term, FGB will rely on the VU Library to provide access to the data package. The VU Library shall consult the relevant department head(s) prior to releasing the data package to an inquiring third party.</p>
<p>3. Guidelines concerning the storage of research data and documentation:</p> <p>3.1 Minimum storage period</p> <p>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.</p>	<p>3.1</p> <p>FGB agrees with the distinction between data and the supporting materials (documentation, metadata, research code/software etc.) with regards to the storage period. Due to the variety of research conducted at FGB, further detail on these storage periods is required. With regards to data:</p>

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^v. 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.

- Specific information on the handling of personal data is discussed in [section 3.2](#)
- If a research project is not subject to the WMO law, the GCP guidelines or the other medical research regulations cited in [section 1.1](#), then a data package created after a research article is published must be archived for a minimum of 10 years from the date of publication (in accordance with VU and FGB RDM policies). If a data package is created upon completion of a research project, but data were not used in any publications (see [section 1.1](#)), the lead researcher responsible for the data package should determine an appropriate duration for archiving. If raw data are archived upon completion of data collection to ensure the secure storage of raw data (see [section 1.1](#)), the archiving duration depends on what is subsequently done with the data. The raw data should, of course, be maintained during the entirety of the research project and if the data are used in any research publications, then the archiving duration should be a minimum of 10 years from the date that the most recent research article was published.
- For research on medical devices, the data package must be archived for a minimum of 10 years after the research project is complete; if the device is released to the market, all data and supporting information about the device must be saved for 10 years from the moment that the last device is placed on the market. If the device is implantable, the term in both cases must be 15 years.
- For research on medicinal products, the following archiving terms must be followed:
 - A minimum of 25 years from the end of the study if the GCP [Regulation 536/2014](#) applies
 - A minimum of 30 years from the expiry date of the product if the [Guidelines on Good Manufacturing Practice Specific to Advanced Therapy Medicinal Products](#) applyAdditional detail on archiving terms for research on medicinal products is found under section 6.3 of the [Guideline on the Content, Management and Archiving of the Clinical Trial Master File](#). If uncertain about which legislation applies, contact research.data.fgb@vu.nl for advice.
- For all other research subject to the WMO law, the data package must be archived for a minimum of 15 years after completion of the research project.
- In some cases, data are reused for new purposes and the results are published in a new research article. In such a case, the storage

duration is extended for another 10 years from the publication date of the new research article.

- Once the storage term is complete and if there is no intention to [publish](#) the research data, the research data should be destroyed, as well as any related personal data that was also archived for the same storage term (see [section 3.2](#) for more information).

NB: In the case of conflict between any of the above archiving term requirements, ensure that the longest minimum requirement is used.

With regards to supporting materials/documentation:

- If there are no intellectual property restrictions on these materials, and it has been confirmed that no personal data from research participants has been included in any of the documentation or software, these supporting materials can be archived indefinitely. Even after the data are no longer available, the persistent storage of the supporting materials supports transparency and research integrity, as well as allowing other researchers to conduct replication studies based on the earlier work.

Lastly, if FGB researchers are confident that the data that will be included in their data package are anonymous and are not subject to intellectual property restrictions, these data may be archived indefinitely. FGB researchers **must** discuss these issues with relevant experts at FGB (research.data.fgb@vu.nl) and VU (rdm@vu.nl) before assuming that the data package may be stored indefinitely. Further guidance can also be found on this page about [privacy risks and anonymous data](#) and this [de-identification guide](#).

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 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

3.2

FGB complies with this requirement with some deviations and further specifications.

For research data that are subject to the WMO law, the GCP guidelines and/or the other medical research regulations cited in [section 1.1](#), these data must be re-identifiable for the entire archiving term to allow for safety monitoring and long-term follow-up. The [directly identifying](#) personal data should be stored in a highly secure archive and kept separate, where possible, from the other research data (see [section 3.3](#)).

For all other research, the FGB researcher(s) should determine if it is necessary to re-identify research subjects for the purposes of research integrity (e.g. in order to confirm with a research subject that the content of their interview data is accurate). If this is not necessary for research integrity purposes, the link between the [directly identifying](#) personal data and the

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.

research data can be deleted, and if the personal data are no longer required, they should also be deleted. However, if the personal data in question represent informed consent data, this data must be saved for the same duration as the research data even after the link between the personal and the research data is deleted. This is because informed consent data are required to demonstrate that consent was legally obtained from participants but the archived research data do not need to be re-identifiable to achieve this purpose. In some cases, informed consent may have been obtained without collecting [directly identifying](#) personal data. In such a case, the consent data must still be stored for the same duration as the archived research data.

Informed consent data containing personal information must be saved in a highly secure archive, preferably separate from the research data (see [section 3.3](#)).

NB: If researchers plan to continue to use the research data for new research questions, and consent was obtained for this purpose, the link between the informed consent data and the research data should **not** be deleted.

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.

3.3

FGB complies with these requirements, with some additional explanation and deviations.

- Firstly, digital archiving storage is provided at a VU-level rather than at the level of FGB and there are different options available for different purposes. Paper archiving is provided at the faculty level on a departmental basis. FGB researchers are expected to use an archive that sufficiently protects their data based on the privacy & security risks posed by the data, even after de-identification. Data stewards from FGB and VU Amsterdam University Library provide support to FGB researchers in determining which archive to choose. The recommendation in the national guidelines to archive raw data in a read-only format before they are made available for processing is essentially the same concept as archiving raw data upon the completion of data collection (see the reasons for archiving in [section 1.1](#)); the goal is to ensure that raw data are fixed which prevents tampering of the data and therefore contributes to research integrity.
- FGB complies with this requirement where possible. As discussed in [sections 1.3](#) and [2.1.1 subsection 4](#), it is often impossible to fully de-identify raw data without irrevocably changing the content of that data. In such a case, the de-identified data still needs to be stored in an archive with a higher level of security. Additionally, because VU Amsterdam can only offer so many archiving options, the raw data that has been de-identified as much as possible

- 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.

and the [directly identifying](#) personal data may need to be stored in the same archive. FGB researchers can consider submitting two related data packages to the same archive, one with the directly identifying personal data and one with everything else, while ensuring that in the documentation for both data packages there is a cross-reference to the location of the other data package. FGB researchers can also consider encrypting the directly identifying personal data, however they should discuss this with a VU Amsterdam IT security expert first to determine how to manage the encryption and de-encryption of the archived data during the entire archiving term.

- If data are archived externally in a national or international archive, FGB does *not* require that the data also be archived locally at VU Amsterdam. FGB only requires that there is sufficient information documented in the data package about where the externally archived data can be found (see [sections 2.1, 2.1.1 subsection 4](#) and [2.3](#)). FGB researchers are expected to check that they are allowed to archive data externally, mainly with regards to privacy concerns.

- FGB agrees that individual storage is insufficient for archiving purposes. FGB researchers are required to use an archive offered by VU Amsterdam or, where appropriate, an external archive which, ideally, should have CoreTrustSeal certification and be, where possible, discipline-specific.

All FGB researchers, not just PhD candidates or postdocs, are expected to maintain a level of documentation and file management that will ensure that another researcher can assume the original researcher's tasks, regardless of whether the original researcher is employed temporarily. To meet this requirement, FGB researchers are expected to prepare their research data management plans with archiving in mind. This means that all of the materials that are required for the data package are prepared and updated throughout the research process and that these materials are maintained in VU approved storage options with regular back-ups to avoid data loss.

- FGB complies with this requirement. See [section 2.1.1 subsection 4](#).

4. Faculty-specific policy

4.

Of the rules listed, the FGB will apply the following:

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.

1. Bachelor's and one-year Master's students should prepare informal data packages, whenever possible. See [section 1.3](#) for details.
2. The FGB recommends, but does not require, that all research data be archived. See [section 1.1](#) for details.
3. Data ownership is determined at the start of a project and defined in the research data management plan. Within FGB, data archiving will not change this pre-defined ownership. If the first author archives a dataset that is owned by another institution, they must include this ownership information in documentation of the data package.
4. FGB may carry out random inspections as suggested by the national guidelines.
5. FGB researchers are expected to prepare and store the data package at the time of acceptance of the manuscript. See [section 2.2](#) for details.
6. FGB researchers are expected to state in the research publication manuscript where data are archived and who to contact regarding questions about the data.

Finally, all VU researchers are required to register a project-level description (also known as project-level metadata) about their archived datasets in PURE, OSF, DataverseNL or YODA.

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

ⁱⁱ 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).

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

^{iv} 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>

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

^{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.