

Guidelines for Archiving of Academic Research for Faculties of Behavioural and Social Sciences in the Netherlands (Version 2.3)

As accepted by the Deans of Social Sciences in the Netherlands, April 2018

Including local guidelines for implementation as laid down by the Faculty of Behavioural and Movement Sciences

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.

Additional note

In the following sections, the guidelines determined at the national level are presented on the left, while the local implementation of the guidelines at the Faculty of Behavioural and Human Movement Sciences (FGB) at the Vrije Universiteit is presented on the right and is based on the principle "apply or explain". The FGB complies with the national guidelines, except on some aspects, which are discussed below.

Research Data Officer FGB: E. Jessica Hrudey, research.data.fgb@vu.nl

National Guidelines	Local Implementation (FGB)
<p>1.1 Purpose of these guidelines These <i>guidelines for the archiving of academic research</i> 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 2018.¹ 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.</p>	<p>1.1 The FGB complies with this requirement, however archiving may be appropriate on more occasions than simply after publication.</p> <p>Data must always be archived after a research article is accepted for publication. Other situations where archiving may be appropriate include:</p> <ul style="list-style-type: none"> • 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 prevent loss or modification of the raw data <p>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.</p> <p>These guidelines only address medium-term data archiving, which facilitates research verification and replicability as required by the VSNU Code of Conduct and the legal requirements for WMO and Good Clinical Practice (GCP) studies. Long-term archival, also known as publishing data, is not applicable under these guidelines and is instead addressed in Annex 3 and in the FGB Research Data Management Policy.</p>
<p>1.2 Existing frameworks within which these guidelines will function</p> <ul style="list-style-type: none"> ○ 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: <i>Wet Medisch-wetenschappelijk Onderzoek met mensen</i>), 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 	<p>1.2 At the moment of implementation of these guidelines, the Personal Data Protection Act (WBP) is no longer valid, and instead the requirements of the General Data Protection Regulation (GDPR) and the Dutch national implementation of the GDPR (UAVG) are applicable to the application of these guidelines.</p> <p>This document is also informed by the references cited at the end of the FGB Research Data Management Policy, as well as the following regulations regarding the Good Clinical Practice (GCP) guidelines:</p> <ul style="list-style-type: none"> ○ Guideline on GCP compliance in relation to trial master file (paper and/or electronic) for content, management, archiving, audit and inspection of clinical trials ○ 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 ○ Commission Directive 2005/28/EC of 8 April 2005 laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorization of the manufacturing and importation of such products ○ Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use
<p>1.3 To whom do these guidelines apply?</p>	<p>1.3 The FGB complies with this requirement. However, it is recommended to have Bachelor's and one-year Master's</p>

<p>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 <i>The Netherlands Code of Conduct for Scientific Practice</i>.¹¹ These guidelines are a concrete embodiment of the 'verifiability' aspect set out in the <i>VSNU Code of Conduct</i>.</p>	<p>students produce an informal archiving package to their research supervisors. This document does not need to be nearly as extensive as a formal archiving package, but should contain the datasets (raw and final), code/syntaxes used, any other documentation on data processing and analyses, and the final research paper.</p>
<p>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'.</p>	<p>2. The FGB refers to the required documentation as an 'archiving package' because archiving may be required in more instances than just after publication.</p> <p>Note that the comprehensiveness of the archiving package will vary depending on the stage of the research project. The FGB is aware that different projects may be at different stages when archiving data, for example with long-term cohort studies or registries. Complex situations are addressed in Annex 1.</p> <p>During the implementation phase of these guidelines, researchers are allowed to submit the required documents in Dutch, if all documentation was completed in Dutch. Starting Sept 1, 2019, all documentation in archiving packages must be in English. This excludes interview transcripts, questionnaires or any other documents that were carried out with study participants in Dutch, or any another language. Where appropriate an English translation of such documents should be included in the archiving package.</p> <p>NB: At the moment of writing of this guideline, the VU policy on the digitalization of paper documents is 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. The VU will decide sometime in 2019 as to whether a validated method for digitalization of paper documentation is feasible and appropriate.</p> <p>NB2: For all studies falling under the purview of the GCP guidelines and/or the WMO, original paper documentation MUST NOT BE DESTROYED until the archiving term is complete, even if a digital copy is made. The Inspection Agency for Healthcare and Youth does not allow the destruction of paper medical charts or informed consent forms until the archiving term is complete.</p>
<p>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.):</p> <p>2.1.1 The published (or accepted) manuscript or publication.</p>	<p>2.1 See Annex 1, for specific information about what is expected within the FGB, including complex situations, such as registry data or ongoing cohort studies.</p> <p>2.1.1 This requirement depends on the reason for archiving. If there is no associated publication, the research protocol should be submitted.</p>

2.1.2. 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.

2.1.3. 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.

2.1.4 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.

2.1.5 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.

2.1.6 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.

2.1.7 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.

2.1.2 This information does not need to be submitted separately in the archiving package if it is clearly described in the publication or protocol.

2.1.3 This information does not need to be submitted separately in the archiving package if it is clearly described in the publication or protocol. If documents, such as questionnaires, were used in multiple languages in the study, all versions used should be submitted.

2.1.4 The FGB deviates from the requirement for providing *anonymized* raw data files. Under the new GDPR legislation, it is very difficult to sufficiently anonymize personal data and too many manipulations are required to achieve some level of anonymity in raw data, resulting in data that are no longer raw. Some variables can be removed from the data to improve anonymity, but this depends on the type of data involved. See Annex 1 and Annex 2 for more information.

Note: in some situations, the data are not in the possession of the researcher who is creating the archiving package (e.g. research with business data that cannot leave the business' servers). If the researcher is unable to store the data in a VU archive, an agreement should be made with the data owner 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 access to the archived data should verification or replication of the research results be necessary. The research can refer to the location of the stored data in the archiving package.

2.1.5 A non-proprietary copy of all code must be provided. Many programs have codes/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. See Annex 4 for more information.

2.1.6 Any intermediate files created during the processing of raw data into analysis 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 the code).

If the raw and/or analysed files are in a proprietary format (such as SPSS, SAS, STATA etc.), it is strongly recommended to archive a non-proprietary (.csv, .tab) copy of the data files, whenever possible.

2.1.7 See section 2.1.5 above

2.1.8 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:

- Name of the person who stored the documents or files
- Division of roles among authors, indicating at least who analysed the data
- Date on which the manuscript was accepted, including reference
- Date/period of data collection
- Names of people who collected the data
- If relevant: addresses of field locations where data were collected and contact persons (if any)
- Whether or not an ethical assessment took place before the research, and, if relevant, statements made by the Ethics Review Committee.

2.1.9 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.

2.1.10 If available, documents relating to the ethical approval.

2.1.8 The data files must be described in terms of their structure (wide files vs long files), the number of records (rows) and variables (columns) in each data file, and how the various files are related to each other (if more than one was utilized for the publication).

The readme/meta-data file should be saved in a non-proprietary format, such as .txt or .xml.

2.1.9 If the researcher intends to publish the research data, the documentation under sections 2.1.2, 2.1.3 and 2.1.8 can be summarized into one readme metadata document.

2.1.10 Upon completion of a research project it is also necessary to archive the informed consent documentation. This documentation should be stored in a high-security archive and, when possible, kept separate from the research data (See Annex 2).

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*).

2.2: FGB researchers are expected to prepare and submit an archiving package upon receipt of acceptance of a research article/publication.

When archiving research upon completion of a project, the archiving package should be submitted within one month of completion. For particularly complex projects (spanning 10+ years and/or with 10+ different data sources), archiving packages may be submitted up to three months after completion of the project.

When raw data are archived after data collection is complete, the researcher determines when archiving should occur.

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

2.3: Within the FGB, this individual responsible depends on the reason for archiving. If the reason for data archival is completion of a project, the lead researcher for the project (e.g. project coordinator, primary investigator and/or project leader) is responsible for the archiving package. He or she may delegate the completion of the package to the researcher most directly responsible for the project, but must review the package prior to submission.

If data are being archived in relation to a publication, the lead author is responsible for creating the archiving package. If the lead author is not an FGB staff member, the FGB co-authors are responsible for checking that the archiving package meets the standards of these guidelines, whether or not the first author works at a Faculty of Behavioural and Social Sciences. If the lead author is not an

<p>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.</p> <p>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.</p>	<p>FGB staff member and does not have adequate data archiving options at his or her facility, the FGB co-authors should work together with the lead author to find an appropriate archiving solution.</p> <p>Supervisors of PhD candidates and Research Master's students are allowed to delegate the preparation of an archiving package to the candidate/student. However, the supervisor should inform the candidate/student about these guidelines and the supporting documents at the start of the PhD/Master's project, so that the candidate can appropriately prepare the archiving package upon completion of the project. As required by the national guidelines, the supervisor is responsible for ensuring the completeness and appropriate storage of the archiving package.</p>
<p>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.</p>	<p>2.4 The FGB deviates from this requirement because of the frequent turnover of research staff within any research institution. In all cases, the department head, and where applicable the department data manager, must have reading rights to the archiving package.</p> <p>If data are archived in association with a publication, the first author must also have reading rights. If the first author ends their employment at the faculty where the data were archived before the archiving term is complete, reading rights must be transferred to the lead researcher for the project.</p> <p>If data are archived upon completion of a project, the lead researcher for the project may also have reading rights, if he or she is still employed at the faculty where the data were archived.</p>
<p>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.</p>	<p>2.5 Duration of research data storage is discussed under section 3.1.</p>
<p>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:</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 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. <p>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</p>	<p>3. The FGB deviates from this requirement because the guidelines cite the obsolete WBP legislation. Within the GDPR any data that can be linked to an individual is considered personal and non-anonymous data. It is therefore not possible, in most cases, to render raw data anonymous. See section 2.1.4 and Annex 1 and 2 for more information.</p>

<p>details, but also photographs and video material. The raw data and the personal data together form the research data.</p>	
<p>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</p> <p>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: <i>Wet op de geneeskundige behandelingsovereenkomst</i>).</p>	<p>3.1. The FGB deviates on this requirement because of the new standards of “anonymity” according to the GDPR. Whether data are raw or processed, if researchers are confident that the data are anonymous, they may be archived indefinitely.</p> <p>Archiving packages that do not fall under the purview of the WMO or Good Clinical Practice (GCP) guidelines, should be archived for a minimum of 10 years.</p> <p>Archiving packages that fall under the purview of the WMO, but not GCP must be archived for a minimum of 15 years.</p> <p>Archiving packages that fall under the purview of the GCP must be stored for a minimum of 20 years (if the Directive 2001/20 EC applies) or 25 years (if the Regulation 536/2014 applies). If uncertain about which legislation applies, contact the research data officer for advice.</p>
<p>3.2 Maximum storage period for research data Data that can be traced back to individuals may in principle <i>not</i> 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.</p> <p>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.</p> <p>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; <i>not</i> in the publication package.</p>	<p>3.2. The FGB deviates on this requirement because of the new GDPR legislation. According to the GDPR, if informed consent is the legal basis for processing personal data, then researchers must be able to demonstrate that consent was obtained legally. Therefore, informed consent forms must also be archived for the durations described under section 3.1. In all cases, informed consent forms must be stored in a highly-secure archive, and when possible, kept separate from the research data. For projects that do not fall under the purview of the WMO or GCP, the key file or ID number that links the informed consent data to the research data should be destroyed whenever possible, in order to reduce the risks of re-identification of research subjects.</p> <p>Once the archiving term is complete, and if there is no intention to publish the data, the data should be destroyed; however, all supporting documentation should be maintained (i.e. the study protocol, publications, metadata and code). This information is still useful to other researchers wishing to start new studies based on previous research.</p>
<p>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.</p>	<p>3.3. See Annex 2 for information regarding the selection of an appropriate archiving system.</p> <p>All researchers 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 or not. To meet this requirement, FGB researchers are expected to prepare their research data management plans with archiving in mind. This means</p>

<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>	<p>that the documents required in the archiving package are prepared and updated throughout the research lifecycle, and maintained in VU approved storage options with regular back-ups to avoid data loss.</p> <p>If any data used in the research project were obtained from external organizations, FGB researchers are expected to appropriately cite these data in the archiving package, as well as include documentation (or citations thereof) regarding how these data were processed by the external organization.</p> <p>The FGB deviates from the requirement that if data are archived externally that “this does not relieve researchers of their duty to store the data internally on the first faculty server”. If a research project is complete and data are to be archived with an external archive, it is not necessary to also archive the data on a local faculty server.</p>
<p>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.</p>	<p>3.4 The FGB complies with this requirement. See section 2.3 above for specifics.</p>
<p>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:</p> <ul style="list-style-type: none"> (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. 	<p>4. Of the rules listed, the FGB will apply the following:</p> <ul style="list-style-type: none"> (i) Bachelor’s and one-year Master’s students should prepare informal archiving packages, whenever possible. See section 1.3 for details. (ii) The FGB recommends that all research data be stored. See section 1.1 (iii) Data ownership is determined at the start of a project and defined in the research data management plan. Within the FGB, data archival 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 the archiving package. (iv) The FGB will carry out random inspections, but only after researchers have had sufficient time to learn and meet the requirements of these guidelines. The first inspections will happen no earlier than Sept 1 2019. (v) FGB researchers are expected to prepare and store the archiving package at the time of acceptance of the manuscript. See section 2.2. (vi) The FGB will phase in these guidelines. See section 2 regarding the phased requirements of all documentation in English. From February 1, 2019, all publications and

(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.

projects that will appear in the 2019 annual academic report should have an archiving package submitted by Sept 1, 2019. After Sept 1, 2019 all researchers are expected to meet the timing requirements discussed in section 2.2.
(vii) FGB researchers are expected to state in the research publication manuscript where data are archived (as well as the DOI if the data are published). This is expected in all publications after Sept 1, 2019.

As stated in the FGB RDM policy, researchers should document all archived data in PURE. The VU Library is developing systems to facilitate this process.

Note: the following footnotes are copied directly from the Guidelines for Archiving of Academic Research for Faculties of Behavioural and Social Sciences in the Netherlands (Version 2.1). Some of the citations describe information from the Personal Data Protection Act (WBP), which is obsolete as of May 25 2018. The footnotes have been kept in this document for **information purposes only**.

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).

ii http://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.

Annex 1. FGB Requirements for Archiving Packages

The requirements for archiving packages within the FGB depend on three factors: 1) reason for archiving, 2) the stage of the research lifecycle at the time of archiving and 3) the nature of the research project (e.g. a one-year study with only a handful of publications expected vs. an ongoing registry or cohort study with no planned end date). In the following table, the requirements for an archiving package (as required by the DSW guidelines, as well as what is additionally necessary within the FGB) are described with consideration for these three factors.

Archiving Package Component	Requirements based on research characteristics
Published manuscript/publication	<ul style="list-style-type: none"> • Required when archiving after publication, regardless of stage of research and nature of the project. • If archiving data upon completion of a project, submit the study protocol and include a citation list of the publications completed with the study data up to point of archiving. • If archiving raw data upon completion of data collection, no publication is necessary. A study protocol should be submitted with the archived data, whenever possible.
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.	<ul style="list-style-type: none"> • Not required as long as the manuscript or study protocol provide this information in detail. • Also not required if previously archived for an ongoing research project and no changes have been made. If previously archived, simply cite the previous location of this information. If changes have been made, submit the new version, including a version number and date to differentiate from the previous version. • If planning on publishing the research data, include this information in the readme.txt metadata document.
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.	<ul style="list-style-type: none"> • Not required as long as the manuscript or study protocol provide this information in detail. • Also not required if this information was previously archived for an ongoing research project and no changes have since been made. If previously archived, simply cite the previous location of this information. If changes have been made, submit the new version, including a version number and date to differentiate from the previous version. • If surveys were used and translated into various languages, all language versions should be included, as well as references to any methods to validate these translations (if this is not already mentioned in the protocol and/or publication). • Additional information that should be included in an FGB archiving package (where applicable to your research and if not already described elsewhere): <ul style="list-style-type: none"> ○ non-response rate ○ timing and location of data collection ○ if applicable, the fieldwork report ○ if applicable, logbooks or lab journals ○ descriptions of measurement instruments used: the device, the brand and version of the device, the conditions during which it was used, calibrations applied ○ hardware and software used (include version) ○ data resolution ○ units used in continuous measurements ○ definitions of categorical variables, and labels for all variables ○ description of quality assurance: checking for errors, validation methods, calibration procedures etc. ○ if applicable, weighting variables ○ any changes to methodology or measurement methods over the course of a longitudinal study ○ descriptions of how derived variables were created (unless clearly described in processing syntaxes) ○ if applicable, information on anonymization methods • If planning on publishing the research data, include the above information in the readme.txt metadata document.

Raw data files	<ul style="list-style-type: none"> • For any project where static (unchanging) raw data files are used for multiple publications, a citation of where the archived raw data can be found is sufficient when creating a new archiving package. • For research projects where raw data files constantly change (e.g registries or long-term cohort studies), the specific raw data file used for the analysis and research publication should be submitted with the archiving package. This raw data file must be clearly labelled with the date and version of the file. Any publications based off of a specific raw data file that has already been archived, may cite it rather than submitting a new copy of this data file.
Analysed data files	<ul style="list-style-type: none"> • If data are archived upon completion of a project, and no analyses were conducted on these data, no analysis file is required. • It is important that the correct version of the file is submitted. To achieve this requirement, researchers should describe effective version control methods for their analysis datasets when writing the research data management plan.
Computer code	<ul style="list-style-type: none"> • All applicable code should be included in the archiving package. In long-term studies, processing code may already be archived. The researcher can cite the location of previously archived code in the new archiving package rather than archiving the code again. • Version control parameters are also necessary for code (both processing and analysis code), particularly in long-term studies and registry data, due to the ongoing changes over the course of the project. The correct version of the code must be submitted with the archiving package and a plan for effective version control of processing and analysis code should be described in the research data management plan.

Annex 2. Choosing an Appropriate Archive

An important step in preparing your data for archiving is to determine whether the information is sensitive because this will inform where you archive your data. Data may be sensitive if there are risks to the privacy of study participants or if there is confidential business information and/or intellectual property information present in the datasets. Under the GDPR, if research data can be traced back to a unique individual, either directly from the research data or in combination with other publicly available information, then these data are still subject to the GDPR. In general, data about humans is rarely anonymous under the GDPR. Even if information such as name, address and contact information have been removed from a dataset, the research data may still not be considered anonymous under EU law. The FGB researcher is responsible for determining the sensitivity of the data; the [research data officer](#) is available for advice when necessary.

Currently the VU offers three archiving options, but it is important to be aware that the VU is working on improving the user-friendliness and storage costs associated with archiving, **which means that the current advice may change in the near future**. At the moment, DarkStor is recommended for archiving all data that are subject to the GDPR. ArchStor can be used to archive data that are not privacy-sensitive, but may contain other sensitive information such as business secrets. DataverseNL may be used to archive all other data. Data recorded on paper can be archived by your FGB department. Contact the departmental secretary for access to this paper archive.

One concern with DarkStor is that data stored in this archive are not quickly or readily accessible to the researcher. Therefore, DarkStor is best used for research projects that have been completed. For data that is subject to the GDPR, but that still needs to be readily accessible to the researcher, the FGB offers a read-only storage option in SciStor. This is not a true archive, but the data stored here cannot be modified which prevents unauthorized changes to the raw data while the research project is still underway. Once a project is complete the data should be stored in a proper archiving solution.

Once again it is important to note that this advice about where to archive data may soon change. Watch out for updates on VUnet and in newsletters regarding new archiving solutions.

Annex 3. Publishing Research Data “FAIR-ly”

Researchers are increasingly expected to publish their research data so that it can be shared with and reused by others. In some cases, research funders require data publication. Within the FGB, the FAIR data principles are gradually being adopted, which means that FGB researchers are expected, **where appropriate**, to publish their research data in a manner that is [findable, accessible, interoperable and reusable](#). Not all data are appropriate for publication, however FGB researchers are expected to assess their data and decide whether they should be published. Reasons for publishing data are discussed on the VUnet page [Research Data Archival](#) under “Data archival: medium or long term?” as well as in the [CESSDA Expert Tour Guide on Data Management](#). If an FGB researcher decides to publish his or her data, the following tips will guide him or her in the process.

Note: it is important to decide early on in a research project whether data will eventually be published and these decisions should be documented in a research data management plan. This will guide how you document and describe your research and, if you are working with data that is subject to the GDPR (see Annex 2), help to ensure that you to obtain consent for data publication from participants while obtaining consent to participate in the research project. You can always obtain consent for data publication later on, however this can be MUCH more difficult. See the [GDPR checklist for informed consent forms and information letters](#) for more information on obtaining consent for data publication. If you are certain that the data are anonymous (see Annex 2), participant consent is not legally required prior to publication of data.

- Currently, the VU does not have any processing agreements in place for data subject to the GDPR with any third-party data repositories. Therefore, for data that is subject to the GDPR, it is not possible to store the data in a third-party repository. **The VU is working on solutions and once a solution is available, researchers will be informed.** At the moment, the best solution is to archive the data in DarkStor and register the archived data in PURE so that the archived data are findable.
 - Published data that are not subject to the GDPR should also be registered in PURE to improve the findability of the data and to demonstrate the data output created by FGB and VU researchers. The VU library is working on solutions to facilitate this process to reduce the burden on researchers.
- FGB researchers must determine how access to the data is managed and controlled. For completely anonymous data with no sensitive business data, the open access option should be chosen. For data that fall under the purview of the GDPR or that contain other sensitive information, the data should only be made available upon request. When data can only be made available upon request, it is important to determine who will be responsible for managing access requests, what is required by the requesting party to gain access, how the data will be safely transferred to the requesting party and what the terms of use are that the requesting party must follow.
 - The FGB, VU library and the National Coordination Point on RDM are developing methods to assist these processes.
 - [IXA](#) can assist with setting up terms of use and data transfer agreements with third parties.
 - If uncertain whether your data should be fully open access or restricted, contact the FGB [research data officer](#) for assistance.
- FGB researchers should use the Dublin Core metadata standards as a default for describing their data (see Annex 4); be aware that research funders may require researchers to use a specific standard and also that certain disciplines have specific standards (see Annex 4).
- In general, the information required for an archiving package can also be used to create metadata for a published dataset. The metadata from a published dataset should at a minimum consist of the information listed in sections 2.1.2, 2.1.3 and 2.1.8 above (see Annex 1 for additional FGB requirements). The information provided should allow another researcher to be able to reuse the data without requiring any other additional information from the original researcher.
- Reuse of data is also dependent on [licensing agreements](#). Licensing of data should be determined early on in the research project, so that all parties involved agree about the reuse and ownership of the data.

Annex 4. Standards for Syntaxes/Scripts/Code and Metadata

Syntaxes, Scripts and Code

Syntaxes, scripts and code (for simplicity, henceforth called syntaxes) must be understandable to other users, whether you are only archiving your data or also planning to publish your data. By appropriately documenting your syntaxes during your research, they are also easier to understand when you need to go back and make changes. Minimum standards are:

- Name and description of the syntax
- Date of creation and any modification dates
- Author(s) of the syntax
- The statistical package and version number used
- Make sure your syntax has a logical flow and break it up into sections to make it easier to follow
 - Make sure to describe what is going on in these sections and any sub-sections. Include references to external documentation (e.g. standards used) if necessary
 - If the syntax becomes excessively long, break it up into separate syntaxes; these separations should be logical: Syntax 1 carries out a certain group of related functions, Syntax 2 another group of related functions
- Use whitespace
 - Avoid stacking functions all directly on top of each other; make space between each function
 - Make use of indentation. Many text editors automatically apply indentation; in SPSS you must actively select this on the task bar
- Save syntaxes and code with descriptive names, version number (if appropriate) and date
 - Some good tips for naming syntaxes (and other files) can be found on the [CESSDA Expert Tour Guide on Data Management](#) website
- Another user should be able to independently use the syntax/code
 - Include coding to open and save data files, so that it is clear which dataset is being processed/analysed and which dataset is ultimately created
 - **Record everything in your syntax or code**; if you absolutely must do something by hand, make sure that it is recorded in your research documentation, and reference this documentation in the syntax

Metadata

There are many metadata standards available for a variety of research disciplines. Applying these standards makes datasets “FAIR-er”: documenting the information required by a given standard makes data easier to find; standards ensure that information about access rights is provided; using standards improves interoperability by only allowing specific vocabularies to appear in the metadata and by giving metadata structure, both of which allow metadata to be read by machines; and other users will be able to understand and reuse the data if sufficient specific information is provided. Metadata standards should absolutely be applied to published datasets; it is also recommended to apply metadata standards to archived datasets.

Standards vary by discipline, but there can also be overlap within research projects (e.g. NTR has both survey data which is covered by DDI and genetic data which would be covered by Genome Metadata). There are also many generic metadata standards such as CERIF and Dublin Core. Many data repositories will generate metadata for you when you upload your data, and the FGB and VU are developing tools to support researchers in the creation of effective and FAIR metadata, therefore knowing all the ins and outs of metadata generation may not be necessary. However, you often need to choose a metadata standard when applying for research funding. Additionally, during a research project, researchers should maintain clear, accurate and organized documentation about the research project (see annex 1), and metadata standards can be used as a guide for what needs to be documented and how. Several metadata standards can be found on the [Digital Curation Centre](#) website. Many of these standards also provide tools with which you can generate a machine-readable metadata document. If no specific metadata standard is required or if several options are available, the FGB-default is to use the [Dublin Core Metadata Generator](#). If you wish to learn more about metadata and want to try your hand at creating some machine-readable documents, such as codebooks that are both human and machine-readable, the CESSDA Expert Tour Guide on Data Management provides an excellent [overview](#) about metadata and the tools available for creating it.