Research Data Management Policy

School of Business & Economics,
Vrije Universiteit Amsterdam

Version 2.0: November 2023

(Further updates of this policy document are expected as data management requirements and standards, legal restrictions, and IT infrastructure develop further)

For inquiries about this document, please contact:

SBE Data Steward
School of Business & Economics (SBE)
Vrije Universiteit Amsterdam
rdm.sbe@vu.nl

Key terms and definitions

Data types

Confidential Data: This refers to data which may only be accessed by authorized individuals. Personal data are one type of confidential data, but the term also applies to details about a business and its management, intellectual property, proprietary information etc.

Personal Data: This refers to information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.[1]

Metadata: Metadata is descriptive information about data so that data can be properly understood and reused well into the future. Simply put, it is data about data.

Sensitive Data: This refers to data that needs to be treated with extra security measures to ensure that it is kept safe from unauthorized persons. It includes both confidential and personal data.

Special Categories of Personal Data: are personal data that reveals racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

RDM during research

Data Management Plan: A Data Management Plan (DMP) is a document that describes how researchers will handle data during and after a research project. It includes information on data collection, storage, processing, compliance with ethics and privacy regulations, data security and protection for privacy, sharing, archiving and publishing.[2]

Data Protection Impact Assessment (DPIA): A DPIA is an assessment to identify the risks of processing personal data and to determine whether personal data may be processed and, if so, what safeguards must be put in place to meet legal requirements.

FAIR principles: Guiding principles for achieving high-quality data, as well as the associated metadata, so that the data can be reused by other researchers. FAIR (meta)data should be Findable, Accessible, Interoperable, and Reusable.[3]

Knowledge Security: This is a policy that ensures that international collaborations at the VU occur in the most secure manner possible and prevent any geopolitical, economic or security risks. Prior to research all researchers engaged in international collaborations are required to assess the risks of engaging in such collaborations, this can be achieved by completing the online ethics self-check.

Privacy Registration: All research projects that process personal data are required to register these data processing activities. This can be achieved by (i) completing the mandatory online ethics self-check, (ii) by completing the VU DMP template 2021. v1.4 and (iii) by completing the VU GDPR registration form for research 2021. Please refer to point 7 in this document for a more detailed explanation on which forms are applicable under different circumstances.

[1] Definition taken from the European Union General Data Protection Regulation: https://gdpr.eu/eu-gdpr-personal-data/

[2] DMP templates are available on the VU DMPOnline site: https://dmponline.vu.nl

[3] See http://dx.doi.org/10.1038/sdata.2016.18

RDM after research

Archiving Data: The long-term storage of research data in a manner that prevents modification, loss, damage or obsolescence of those research data. This is encouraged to allow for (i) verification of results and to uphold academic integrity; (ii) the reuse and replication of research results (iii) sustainability of data so that it can be used by others in the long-term.

Data Package: The supporting research project materials, such as README files, data, metadata, replication code and other documentation that need to be archived after a research project is complete and/or a research article based on these materials has been published.

Dataset Registration: This refers to the public sharing of information on the location of your archived or published dataset as well as other related metadata in the VU Research Portal PURE. At SBE, the registration of datasets is an important criterion for the calculation of research time.

Persistent Identifier: A durable reference to a digital dataset document, website or other object. By using a persistent identifier, you make sure that your dataset will be findable well into the future. A DOI or Handle are the commonly used PIDs.

Publishing Data: This refers to the public disclosure of research data. This ensures that data are findable, accessible and reusable by other researchers. Published data do not necessarily have to be made open access, in cases where data cannot be shared for confidentiality purposes,researchers can share the metadata and data documentation and restrict access to the datasets.

Published Research: Research published by SBE affiliated researchers in a publication in a peer-reviewed journal, a book, or a book chapter.

General Introduction

Good data stewardship is one of the prime responsibilities of a professional research organization. Data stewardship implies that data is managed in a professional and careful manner throughout all stages of research projects (i.e., the design, collection, storage, processing, analysis, long-term preservation, and publishing of research data). The Vrije Universiteit Amsterdam (VU) adopted a university wide data management policy, updated in 2020. This document elaborates the faculty-specific data management policies of the School of Business and Economics (SBE), as a specification of, and supplement to, the VU-wide policy.

Data management is a multi-faceted concept. The focal points of the current policies are primarily:

- INTEGRITY: ensuring academic integrity, through transparency and research verifiability;
- SECURITY: ensuring that data are stored and transferred as securely as needed, with minimal risks of data loss, and in compliance with legal requirements and research ethics guidelines, in particular for sensitive data;
- REUSABILITY: stimulating the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) to enhance research progress.

These policies can best be summarized as “as open as possible, as closed as necessary”.

Aims, scope and responsibility

1. The policies regarding data management and data stewardship at SBE as laid out in this document explicate and specify the general research data management policy of the VU[4], which in turn incorporates guidelines from the Netherlands Code of Conduct for Research Integrity.[5]

2. The aims of these policies are to help SBE researchers:

    - meet the SBE Research Ethics Regulations for researchers whose work deals with human subjects and the legal requirements stated in the EU General Data Protection Regulation (GDPR) and the Dutch GDPR Implementation Act (Uitvoeringswet AVG). Specifically, these legal documents provide recommendations on how to protect the privacy of human study subjects when processing personal data;
    - meet the requirements of research funding agencies;[6]
    - meet the requirements of scientific journals concerning quality, availability, and transparency of data and code;[7]
    - protect the faculty's and their own academic integrity;
    - avoid data losses, data breaches, and risks that ensue from international collaborations;[8]
    - stimulate research progress, open science and enhance research quality through the sharing of data, code, and documentation.[9]

3. The policies in this document apply to **all** research aimed at a publication in a peer-reviewed journal, a book, or a book chapter by SBE affiliated researchers.[10] For the remainder of this document, the expression "published research" or "research" is reflected in the definition given above.

    In **all cases** researchers must comply with the legal and ethical requirements stipulated in this document.

4. Each individual SBE researcher is responsible for adhering to the policies in this document and in the VU RDM policy. Researchers with questions can contact the SBE data steward for advice (rdm.sbe@vu.nl). Further support can be obtained by contacting the VU RDM support desk (rdm.support@vu.nl).[11]

5. In line with the VU RDM policy, **department heads** are responsible for ensuring that RDM policies are followed by researchers within their departments; that data are stored securely

[4] https://libguides.vu.nl/rdm/policies-regulations

[5] As updated in 2018: Netherlands Code of Conduct for Research Integrity

[6] See for instance, https://www.nwo.nl/en/research-data-management and https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdfn

[7] The Center for Open Science produces an overview of the transparency and openness requirements of a wide range of journals: https://topfactor.org

[8] https://vu.nl/en/employee/onderwijs-en-onderzoeksbeleid/knowledge-security-for-vu-amsterdam-employees

[9] https://vu.nl/en/about-vu/organisations/open-science

[10] Thus, it covers the entire complex diversity of research cultures and data needs at SBE, such as experimental research, surveys, commercial databases, proprietary internal data, externally managed databases, qualitative and quantitative data, big real-time data sets, etc. Future versions of these policies may extend the scope of application further, for example to published working papers and student research projects. A proviso holds for research that builds on previous research for which the data collection process was initiated before January 1, 2017, and for which not all steps in the current SBE data management policies can be retraced. Here researchers should comply with the policies on documentation and archiving as far as is reasonably possible. Researchers should always comply with legal and ethical constraints (e.g. for sensitive data).

[11] See https://vu.nl/en/employee/research-data-support

using VU recommended storage services; that data is made accessible; and that datasets are archived and registered in Pure and the VU Research Portal.[12]

Additionally, department heads should ensure that data transfer and archiving arrangements are made when a researcher's employment contract comes to an end. Department heads can designate one of their department members to act as a data coordinator who will support the department head and researchers in data management processes.

Before research (planning)

6. A **Data Management Plan** is mandated for all PhD projects at SBE and for research projects whose funding agencies have made this a requirement. For instance, preparing a Data Management Plan is a requirement for projects funded by the Dutch Research Council (NWO) and the European Research Council (ERC). Data Management Plans can be created using DMPOnline (https://dmponline.vu.nl/) which is a tool that offers practical tips and guidance on creating a DMP.

   So far, SBE does not require a Data Management Plan for all research conducted in the faculty, but all empirical research projects are encouraged to prepare one. Data Management Plans are particularly recommended for large projects and new lines of research. PhD candidates can write a Data Management Plan as part of their mandatory Research Integrity (ABRI) and Academic Integrity (Tinbergen Institute) courses. All Data Management Plans created at the SBE have to be shared with the SBE data steward for feedback and privacy registration. Questions related to the preparation of Data Management Plans can also be directed to the SBE data steward (rdm.sbe@vu.nl).

7. **Personal data** warrants special attention, as there are substantial legal and ethical requirements for handling this type of data. The GDPR defines personal data as information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Special categories of personal data include data on race, ethnicity, religion, political opinion, philosophical beliefs, trade union membership, genetic and biometric traits, sexuality, sexual orientation and health. If in doubt whether your data is legally classified as personal data, contact the SBE's privacy champions (privacy.sbe@vu.nl).

   Researchers who collect and process personal data in their research have a legal requirement to register their research projects in the Privacy Registry. This is a legal obligation imposed by General Data Protection Regulation (GDPR). At SBE, the registration of research projects that deal with personal data can be done in three ways:

   - By answering the relevant GDPR questions incorporated in SBE's online ethics self-check tool. This is strongly recommended for all research projects at SBE that deal with data on human subjects as it integrates ethics, GDPR and privacy risk assessments. Research projects that handle personal data and that are required to create a Data Management Plan by their funding agencies (e.g. ERC, NWO and ZonMw) will have to complete both the SBE's online ethics self-check tool and the funder DMP template;
   - By completing the *VU DMP template 2021 v1.4* available in DMPOnline. This template is recommended for all research projects that need to create a Data Management Plan such as PhD projects but are not funded by a funding agency. The GDPR-related questions in the online ethics self-check can be skipped if this template is used;
   - By completing the *VU GDPR registration form for research 2021 v1.1.*

[12] See https://libguides.vu.nl/rdm/dataset-registration

8. The GDPR stipulates that a Data Protection Impact Assessment (DPIA) is required for projects that are likely to impose a high risk on research subject's personal data. To assess the risks associated with using personal data, researchers should complete the risk-assessment section of the online ethics self-check.[13] Researchers that respond affirmatively to at least two of the questions in this section are required to complete a DPIA and should contact the faculty's privacy champions (privacy.sbe@vu.nl) to access the DPIA template.

9. If a VU researcher would like to process personal data or confidential data obtained from other parties (e.g., ministries, corporates, banks, etc.), the researcher should first question whether the privacy legislation (GDPR) allows the VU to collect or receive these data for research. When data is collected by the VU through questionnaires or interviews, the basic principle is that informed consent from the participants is required. When it comes to data that has already been collected, it will have to be examined whether the informed consent obtained by the third party when collecting the data includes secondary use or if the data may be used without consent. In all cases, data sharing agreements will have to be made between the VU and the third party about security, confidentiality, privacy, intellectual property and publications. In addition, the data subjects must be properly informed about the processing of their personal data[14]. The faculty privacy champions can be contacted for support (privacy.sbe@vu.nl).

   For receiving personal data from, collecting personal data in and/or sending personal data to countries outside the EU, additional legal and ethical rules may apply. The researcher is responsible for compliance with such rules and is expected to contact the SBE privacy champions (privacy.sbe@vu.nl) and the SBE Ethics Review Board (rerb.sbe@vu.nl).

10. All researchers that engage in international collaborations are expected to complete the questions in the knowledge security section of the online ethics self-check to assess whether their collaboration will result in potential unethical consequences or national security threats. Researchers should contact the research office (researchoffice.sbe@vu.nl) if they suspect that their international collaboration cannot occur securely or if they have questions about the knowledge security framework at the VU and its implementation at SBE. [15]

During research (data collection and storage)

11. For ongoing research, data should be **stored** securely and professionally, meaning that measures are undertaken to prevent data loss (this includes using back-up facilities and proper hardware maintenance) and data leakage. The loss of a single data carrier or the departure of an individual researcher should not prevent the rest of the research team (at VU or elsewhere) from retrieving the data. Researchers should make sure that back-ups of the data and their accessibility are properly arranged and communicated within the research team. Responsibility for ensuring this lies with the principal researcher of the project.

12. During research, researchers should ensure that it is clear how and where data and code can be **accessed**. This documentation of data storage is particularly important when researchers leave the VU. Details on the storage arrangements should be known to at least two people in the department at any time, one of whom is the department head or the department data coordinator (see point 5). Researchers should make sure that data are accessible for verification

[13] The risk-assessment section is based on the questions in the VU pre-DPIA template that has been prepared by the VU privacy lawyers.

[14] Please note that all legal documents at the VU including data sharing agreements have to be signed by the SBE managing director who is the legal signatory of the faculty and not by the Vu researcher. Researchers are advised to contact the VU legal office (legal@vu.nl) for the review of the agreements and afterwards the SBE secretariate (office.sbe@vu.nl) to have the agreements signed on their behalf.

[15] Refer to https://vu.nl/en/employee/onderwijs-en-onderzoeksbeleid/knowledge-security-for-vu-amsterdam-employees for more information on the VU knowledge security framework.

and reuse. Furthermore, researchers should be able to grant research subjects access to all information gathered about them, to the extent that it has not yet been fully anonymized.

13. For sensitive data (personal data or confidential data) additional security measures are needed to ensure the privacy and interests of research subjects. Researchers are advised to make use of the data classification tool (https://vu.nl/en/research/dataclassification) that distinguishes across four levels of data sensitivity (see Appendix 2 for explanations of these levels). Following data classification, researchers should use the storage finder to select appropriate data storage options (https://vu.nl/en/research/storagefinder). In general, VU recommended storage solutions such as Research Drive and Yoda should be used for the storage of data that has been classified as having low, medium and high sensitivity levels. Research Drive is suitable for data that involves the creation of multiple sub-folders, is dynamic and regularly needs to be updated or has multiple collaborators across various institutions. Yoda is suitable for the storage of data with less complex folder structures and doesn't require frequent updates.
    Data classified as having medium and high sensitivity might require encryption before storage on Research Drive and Yoda. Medium and high sensitivity data should still be encrypted if a researcher uses a PC or laptop with encryption. In addition, PCs should always be locked if a researcher has stored sensitive data and is absent.[16]

    Researchers whose research involves data classified with the highest level of sensitivity should contact the faculty privacy champions as special custom security arrangements will be needed for such data (privacy.sbe@vu.nl). Data that has been classified as having medium, high and the highest levels of sensitivity should never be stored on unprotected data carriers (including unencrypted hard-drives in password protected[17] computers) or on synchronized commercial cloud services (Dropbox, Google Drive, One Drive)[18], particularly if these mirror to local non-encrypted hard-drives.
14. Researchers should, whenever possible, **anonymize** or **pseudonymize** personal data before using, sharing or storing the data. Anonymization and pseudonymization of personal data reduce the risks associated with the use of such data and entails a shift from high or medium sensitivity to low sensitivity levels. Encryption keys (see point 12) and keys linking the pseudonymized data to the personal data should be kept safely and separately and be accessible by at least two persons affiliated with the department at all times, including after the departure of the original researcher. It is advised that next to the researcher, at least the principal researcher or department head and one more person (such as the departmental data coordinator) have access to the keys, and that these keys are stored in at least two restricted locations. The locations for key storage should be either physically (safe) or electronically (at a secure medium) sufficiently well secured and separate from the encrypted, pseudonymized data. The ID number of individuals in the database cannot contain any potentially meaningful identifiers (such as initials, date of birth, postal code). [19] The faculty data steward can be contacted for support with data encryption, anonymization and pseudonymization (rdm.sbe@vu.nl)

16. If a researcher wants to share personal data with outside parties (including co-authors), the researcher should make sure they comply with the legal requirements. It is not always possible to share personal data. The first thing to consider is whether the given informed consent also provides for the sharing of data with other researchers. If this is not the case, the possibilities of sharing the data will have to be examined. An important tip when drawing up informed consent is to mention the possibility of sharing data with other researchers and not to formulate the data

[16] This can be achieved by simultaneously pressing the keys Ctrl + Alt + Delete on the keyboard
[17] The hard drive could be removed and copies to be read on another machine
[18] Exceptions are made for the storage of low and medium sensitivity datasets stored on synchronized cloud services that are GDPR compliant for instance the use of Google Drive with a VU account and GDPR proof Dropbox. The use of these cloud services should be discussed with the SBE Privacy Champions (privacy.sbe@vu.nl).

accessibility too narrowly. If it is possible to share the data, it is important to sign a Data Sharing Agreement with the receiving party. The VU legal office has drafted model agreements for this purpose. A separate agreement is available for researchers who want to share data with students. The researcher should consult with the faculty privacy champions (privacy.sbe@vu.nl) to access the latest version of these agreements and to be fully briefed on the requirements to share personal data outside VU.

17. Researchers that make use of data processing services to process personal data such as survey collection platforms, transcription and data analyzation services need to comply with legal requirements that protect the privacy of personal data. If no agreement exists between the VU and the service provider, a Data Processing Agreement will have to be completed. Researchers should contact the faculty privacy champions (privacy.sbe@vu.nl) to find out whether agreements exist with the preferred service providers, to access the latest version of these agreements and to be fully briefed on the requirements to share personal data with data processors.

18. Researchers should not **travel** with data that has been classified as medium sensitivity, high sensitivity and highest sensitivity data in unencrypted format. In a number of countries importing encrypted data is illegal and should be avoided (e.g., Russia, China). If researchers need to travel with an encrypted laptop to secure their data, they should seek advice from the faculty privacy champions (privacy.sbe@vu.nl)

19. **Big data.** Some datasets are too large for (standard) storage. The service storage, SciStor-Storage for Scientists is recommended for the storage of large datasets. Information on the use of SciStor can be obtained from the faculty's data steward (rdm.sbe@vu.nl) or the research data management support desk (rdm.support@vu.nl l). Researchers handling big datasets that require high performance computing (HPC) can contact IT for Research (itvo.it@vu.nl) to discuss the different HPC options available at the VU. Moreover, researchers are encouraged to use the best practices in their field, and share these explicitly with the faculty's data steward (rdm.sbe@vu.nl) such that the faculty can develop a more concrete policy in this area.

After research (documentation, archiving and publishing)

20. As part of our commitment to open science, we encourage researchers to **publish** data and the corresponding data documentation, in an open manner whenever possible. Publishing data refers to the public disclosure of research data. Ideally, data (including details on the data collection, data processing and steps required for study replication) are published in an open repository (e.g. DataverseNL, Yoda without any accessibility restrictions or the Open Science Framework) such that lasting access of the publication, the data, and the data documentation are ensured, and data may be re-used.[20] Data and documentation can also be published on the journal's website along with the original publication. Please note that personal data should in principle not be made publicly available.

    If open or complete publication of data and documentation are not possible (because of privacy or confidentiality concerns), raw data and documentation need to be securely **archived** for a minimum period of 10 years, also in cases where the individual researcher is no longer available. Archived data should be available for the purposes of research verification whenever academic integrity concerns arise. Archiving data refers to the long-term storage of research data in a manner that prevents modification, loss, damage or obsolescence of those research data. Yoda is recommended for the archiving of datasets with closed or restricted access.

[20] For example, Elsevier has introduced the "Data in brief" option upon acceptance of a paper. Similar initiatives are expected elsewhere.

21. For publishing and archiving data, the procedures for data collection and analysis should be **well documented**, such that they are (in principle) verifiable. The documentation should preferably be part of the published research itself, or of (online supplemental) appendices of the published work. If not, the researcher should provide detailed supplemental README files and archive these securely along with the data in a data package. The data package and documentation should be sufficiently detailed and include:

    - One or more README files describing which documents and files can be found in the package as well as where and how they should be interpreted. A template for such a README file that includes: the names and roles of contributors, date and period of data collection, information on whether an ethics assessment took place has been included in the appendix;
    - details on the process of collection (sampling, selection) of the raw data (including survey data, interviews, video material, experiment scripts, details/code on how websites were scraped, etc., if appropriate);
    - detailed data descriptions, including descriptions of the variables (possibly also with database tickers/acronyms in case external databases are used, and including details on the interpretation (e.g., does 1 indicate male or female as a gender dummy));
    - details on filters and data manipulations used to get from the raw data to the data used for the empirical analysis (including the removal of outliers or individuals from the original sample (e.g., on Trial 5 of Participant 10 there was other student interference and hence the responses are replaced by NaNs), operations on variables (winsorizing, trimming, scaling, rotating), details on how recorded interviews were transcribed and coded, etc.);
    - processed research data along with the codes and scripts used to produce and analyze it. This includes C- codes, Matlab, SPSS, SAS, STATA, R, Eviews scripts, etc., including a README file on the version of the language or package in which the codes were run. [21]
    - Statement from the SBE Ethics Review Board (if needed; see SBE's Research Ethics policies)[22] and evidence of informed consent from research subjects, if applicable;
    - the published research associated with the archived data;
    - details on how privacy issues are dealt with such as anonymization, pseudonymization and encryption (if personal and confidential data are involved);
    - details on how access to the data is arranged and ensured (at least for internal purposes) and who can access the data for the minimum period of 10 years.

22. The quality of the stored and documented data should comply with the standards in the Netherlands Code of Conduct for Research Integrity. This includes the requirements that (i) all steps in the research process can be checked and should (in principle) be replicable, and (ii) that the quality of collection data, data input, data storage, and data processing are monitored and controlled well.

23. After research is published, researchers should **archive** their raw[23] and processed research data, unless compliance issues arise (e.g., license or confidentiality issues), or storability becomes an issue (for some big data sets). Ideally, data should be archived within a month of the definitive publication. Archiving means storing data on a secure system, *together with* the data package

[21] Though not enforced by the current policy document, researchers are encouraged to provide clean, annotated and readable versions of their code.

[22] See https://vu.nl/en/about-vu/faculties/school-of-business-and-economics/more-about/research-sbe

[23] What precisely is labeled as the raw data may be a trade-off between efficiency of storage and verifiability, particularly in qualitative research processes. For example, transcripts of interviews or their coded versions may sometimes be classified as the raw data, rather than the original audio recordings. Similarly, the scraped internet information may be classified as the raw data, rather than snapshots of each of the underlying websites that was scraped.

(see point 19 above). To comply with VU policies, research data produced at or analyzed while affiliated to the VU need to be archived securely for a period of at least 10 years.[24]

Researchers do not need to archive datasets that have already been archived by co-authors with other institutional affiliations if the datasets are archived for a period of ten years and well documented (see point 21). Data that cannot be archived, e.g., due to legal constraints, should still (in principle) be verifiable by archiving the data documentation and publishing the metadata (point 20 above).

Recommended and VU-supported archiving solutions are:

- **DataverseNL**, an online platform for the publication of research data in a semi-open environment. DataverseNL allows users to link directly from publication to dataset and to share it.[25] It is suitable for non-sensitive datasets that are smaller than 2 TB.
- **Open Science Framework** a web-based open-source research project management tool that supports researchers throughout the project lifecycle. It can be used for study pre-registration, data storage and data archiving. It is ideal for non-sensitive datasets that are smaller than 2 TB.[26]
- **Yoda**, is a research data management service that enables researchers to securely store, deposit, share, publish and preserve large amounts of research data during all stages of a research project. Yoda is suitable for the archiving of sensitive datasets as well as for large datasets up to 10 TB.

Other archiving-options may be possible as long as they are certified with a CoreTrustSeal and the data remain available for 10 years.[27][28]

24. Archived and / or published datasets that have been generated at the VU should be made findable (the F in FAIR) by **registering the dataset** in the VU Research Portal (Pure). Dataset registration indicates the nature of the data, the repository used for archiving or publishing the data, a contact person for the data, and the steps needed to request access to the data. Researchers who publish datasets along with a paper or in a discipline specific repository can link those locations to Pure, preferably with a persistent identifier. When the dataset cannot be made publicly available, the description of the dataset should still be registered in Pure. In this case, the reason why the data cannot be published (e.g. confidentiality) should be included in the dataset registration. Datasets should still be registered in Pure if the data is generated by or with co-authors not affiliated to the VU. Along with the registration in Pure, findability can also be achieved by registering the dataset in another discipline-specific repository.[29,30]

25. Data should be archived in a digital format (not in paper format) as much as possible. File should be archived in a sustainable format that is suitable for long-term preservation and accessibility

[24] Data that are relevant for re-use are preferably archived for a longer period. Research data that may be relevant for future research and that may need to be stored longer includes: (1) unique data: Data that cannot be collected again. For example: the operating temperature of today; (2) data of scientific or historical value: research which reflect a period in history, such as an interview with a veteran of World War II or photographs from a certain period; (3) valuable data for re-use: Data that could be valuable to other researchers so that they do not need to collect or structure the data themselves. Such data would ideally be stored indefinitely.

[25] See https://dataverse.nl

[26] see https://osf.vu.nl/

[27] This excludes storage options where the individual researchers pay a fee for the storage facility to continue its services: in this case the storage depends on the availability of the individual researcher for the period of 10 years. Such facilities may only be used if the financial commitment is transferred to the researcher's department.

[28] https://www.coretrustseal.org/why-certification/certified-repositories/

[29] For instructions how to register datasets: see Appendix "Guideline for registering datasets in Pure".

[30] This does not apply to data that is owned by others or where agreements prohibit making any meta-information about the data public.

e.g. .pdf, .txt, .csv, etc.[31] The most sustainable formats should also be used for the archiving of image or movie files, examples of such formats include: .jpeg, .mxf and .mkv.[32] All data files should contain information (or accompanying README files) about the software that is needed to open the file (and which version of the software has been used). Questions on suitable file formats can be directed to the faculty data steward (rdm.sbe@vu.nl).

26. After the research has been completed and / or the data archiving period has elapsed, all data carriers with personal data should be deleted in accordance with legal requirements. Researchers that produce datasets that are difficult or too costly to replicate or related to new lines of research should discuss the possibilities of extending the archiving period with their heads of departments as additional archiving costs may arise.

27. **Secondary data** are data that are collected by for instance a third party for their own research, data that are collected by companies internally, or data collected by institutions that specialize in data collection (such as CBS, Eurostat, FED, Datastream, Reuters, Bloomberg, etc.). If the data provider allows for archiving the secondary data, this is preferred. However, secondary data need not be archived if the data are (in principle) recoverable (possibly after paying a fee or establishing the appropriate contacts). The data documentation (including the process of collection and constructing the raw and cleaned data) should still be archived and contain sufficient information on how the data were obtained and accessed, and details (and possibly scripts) to get from the raw (possibly proprietary or commercial) data to the data used for the empirical analysis, similar as in the standard case. This includes the mentioning of company contacts if the data were obtained through private contacts or if agreements with the company disallow the local archiving of the data at the VU.

28. Data management tools offered by the VU (e.g. Research Drive, Yoda) are free to use up to a certain limit (e.g. 500 GB), beyond which a fee needs to be paid.[33] Researchers, in coordination with their head of department, should secure the relevant funding for this upfront, if needed. Projects funded through a funding agency such as the ERC or NWO may cover the costs of data storage and archiving through the budget set aside for data management costs. Researchers are encouraged to incorporate these costs in budgets of project proposals.

Hardship clause

29. Any exceptions of the above data policies shall be decided upon by the faculty board upon advice of the faculty's scientific committee.

[31] Note that data files of standard software such as SAS, STATA, SPSS, Matlab, etcetera, but also well used formats such as .xls and .xlsx or .docx are less robust as they are not always transferrable from one version of the software to the next. The simplest formats such .txt or .csv should be used.

[32] A list of preferred file formats is available here: https://dans.knaw.nl/en/file-formats/

[33] Please refer to the cost model available here for data archiving and storage costs: https://vu.nl/en/employee/research-data-support/costs-research-en-archiving-storage

Appendix 1 : README FILE PUBLICATION PACKAGE

Publication Information:
Title: *title of publication the publication package belongs to*
DOI: *DOI or URL of the publication*
Date accepted: *Date the manuscript was accepted*
Formatted bibliographic reference: *reference in any format*

Authors
Fill out the table, making sure to note all and contributors.

Name	Affiliation	ORCID	ROLE IN PROJECT

Data collection and processing:
Date/period of data collection:
Names of people who collected the data:
Methods of data collection*: State whether surveys, interviews, or experiments etc. were used to collect the data*
Location of data collection: *(If relevant:) addresses of field locations where data were collected and contact persons (if any)*

Privacy and Ethics assessment:
Describe if ethics assessment took place
Describe measures taken to deal with sensitive data such as encryption, anonymization and pseudonymization

Data availability and license:
Data availability: *Describe the availability of the data here and the ways in which it can be accessed. If data cannot be published, include a valid reason*
Data License: *If applicable state which license applies*

Folder Structure:
Describe the folder structure used in the package and explain which folders include data documentation, raw and processed data

Appendix 2: Data Classification, security, storage and archiving implications

Here we provide a guideline that outlines the different levels of data sensitivity for data at SBE as well as the implications for data security, storage and archiving that arise with the use of data across the different data classifications. Please note that the examples provided are not comprehensive and that researchers are advised to contact the faculty data steward in case of uncertainty. This data classification scheme and recommendations will be updated over time.

Data Sensitivity Type	Definition	Examples	Security Implications	Storage recommendations	Archiving recommendations
Low (public)	Data should be classified as Low (Public) when the unauthorized disclosure, alteration or destruction of that data would result in little or no risk to the University, its affiliates or research subjects.	• Public data of non-vulnerable persons that can be disclosed without any repercussions • Organisational charts and other records that are publicly available	Data does not have to be encrypted before storage	• Research Drive (recommended) • Yoda (recommended) • Open Science Framework (recommended)	• Open Science Framework • Dataverse NL • Yoda
Medium (private)	Data should be classified as Medium (Private) when the unauthorized disclosure, alteration or destruction of that data could result in a moderate level of risk to the University, its affiliates or research subjects.	• Aggregated, anonymised or pseudonymised data where individual respondents cannot be identified • Interviews or questionnaires with non-vulnerable subjects that have given explicit and informed consent to disclose their data (in cases where anonymisation is not possible)	• Data should only be shared with authorised persons within research team. • Informed consent should be obtained before the collection of personal data. • Data should be encrypted before storage and archiving if it cannot be deidentified. • Data should not be stored on local storage, unguarded,	• Research Drive (recommended) • Yoda (recommended) • Open Science Framework (recommended)	• Open Science Framework for de-identified data • Dataverse NL for de-identified data • Yoda for personal data

		• Data classified as ‘for internal use’ by others	unprotected devices or servers that do not use multifactor authentication if it cannot be de-identified.		
High (restricted)	Data should be classified as High (Restricted) when the unauthorized disclosure, alteration or destruction of that data could cause a significant level of risk to the University, its affiliates or research subjects	• Personal data and special categories of personal data of study subjects • Financial information • Data classified as ‘confidential data’ by others • Encryption keys • Datasets with upto a 10,000 subjects (big data)	• Data should only be shared with authorised persons within research team. • Data should be encrypted before storage and archiving • Data should not be stored on local storage, unguarded, unprotected devices or servers that do not use multifactor authentication. • Informed consent should be obtained before the collection of personal data.	• Research Drive with encryption (recommended) • Yoda with encryption (recommended)	Yoda
Highest (secret)	Data should be classified as Highest (Secret) when the unauthorized disclosure, alteration or destruction of that data poses an unacceptable level of risk to the VU, its affiliates or research subjects.	• Medical files of famous (Dutch) people or detainees; • Medical files on hereditary or psychological information • Data classified as ‘top-secret data’ by others • Datasets with personal data collected from more than 10,000 people	Contact the faculty privacy champion during the planning phase of such a project	Contact the faculty privacy champion during the planning phase of such a project	Contact the faculty privacy champion during the planning phase of such a project