Research Data Management-protocol of the Faculty of Religion and Theology of the Vrije Universiteit

Introduction and scope

Data plays a central role in any scientific endeavour. In principle, research is a question-driven, methodical analysis of a specific dataset. In some cases the dataset is fixed, whilst in other cases it is the researcher themselves who generates it. In all cases, research data management (RDM) will play a role. This protocol describes the codes of conduct for employees, external researches and students conducting research that is connected to the Faculty of Religion and Theology at the VU. It aims to answer the question: how to deal with data?

The most important guidelines are those concerning personal data, as this type of data is protected by law and requires the greatest care. For students, it is essential to be aware what working with personal data requires of them. For use of personal data, one has to comply with the General Data Protection Regulation (GDPR, in Dutch: Algemene Verordening Gegevensbescherming or AVG).

Personal data refers to all information relating to a natural person. This includes information such as names and addresses, bank account numbers, IP addresses, and various background information about persons (e.g. race, religion, health condition, political preference, sexual preference, and study progress). Due to the sensitive nature of these data, you cannot process these in the same way as other types of data.

Why a RDM-protocol?

In 2018, the VSNU (Association of Universities in the Netherlands) has determined – in *Gedragscode Wetenschappelijke Integriteit* (Netherlands Code of Conduct for Research Integrity) – that it is one of the duties of a university to take proper care to handle data integrally. The ongoing digitalisation, which also effects the research structure of the FRT VU, prompts researchers to be more aware of the responsibilities of data management.

Data leaks can lead to liability claims and loss of data can result in the failure of a research project. The implementation of this protocol is vital if the FRT VU is to satisfy the duty of scientific integrity. This protocol is an implementation of the broader RDM-protocol of the VU as a whole.

Contents of this protocol

First, a brief description of Research at the FRT VU will be given. Then, the role of the Data management plan will be described. Next, the Data life cycle will be discussed, as well as the principles relevant to the Data storage and archiving. Personal data management will be given special attention due to its sensitive character, and because this is most relevant to students. Some attention will also be given to the themes of Responsibility and ownership of data management, External PhDs, Support (from the VU), and Embedding in the organisation of the FRT. Finally, some *Useful DMP-related information* will be provided.

Research at the FRT VU

This faculty investigates the phenomenon of religion in all its facets. Specifically, the disciplines, practiced at the FRT, can be divided along four methodological lines; research can have a literary, historical, systematic, and/or empirical character. Some research projects can be, exclusively, placed in one of these categories, while other projects use a combination of two or more of these types of methods. Pragmatically, the following distinction can be made: the literary, historical and systematic research focuses on available data, which are subsequently ordered by the researcher, but are – in principle – publicly accessible; empirical research, on the other hand, generates its own data. In this area of research (i.e. the empirical), personal data plays a role – through contact with people, at an individual or group level. Hence, this segment of FRT-research has to be particularly aware of privacy-legislation (GDPR, AVG), as well as ethical restrictions. Although these concerns are an important focus of this protocol, data play a role within all types of research (literary, historical, and systematic, as well as empirical). This protocol, hence, uses the working definition of 'data' as given by Christine L. Borgmans: 'Data are representations of observations, objects or other entities that are used as proof for phenomena within scientific research.'[1]

In practice – within the humanities – a distinction is made between four forms of data:[2]

- Primary or declarative data: the primary (source) material that forms the basis of a project;
- Procedural or derivative data: data gained through analysis of the primary data;
- Metadata: data about the documents containing the derivative data; and
- Operational data: data about the functioning of an IT-system.

Research data– on all of these four levels – can play a role throughout all disciplines (not just as a phenomenon exclusive to social sciences). A historical study, for example, can start from archived material. This archived material is the primary or declarative data. The notes of the researcher, based on the archived material, are the procedural or derivative data. Documents, tracking the progress of the research, are the metadata, and information on the operation of a potential digital application, wherein the data is processed, is the operational data. Hence, data management is not exclusive to empirical researchers, and thus this protocol will not focus exclusively on, e.g., personal data.

Data management plan

Senior researchers will always – notwithstanding the disciplinary embedding of the project – have to compose a data management plan (DMP) for themselves, online the VU offers tools to aid in the composition of DMPs (more on this below). Every project should have its own DMP. This is particularly relevant for (governmentally) subsidized research or externally financed projects. Usually, a DMP will be required for such projects. For students, a DMP is (formally) only relevant for the research proposal of a thesis. However, particularly those students working with personal data for, e.g., their thesis, should take careful note of the following.

The following questions are to be answered in a DMP: [3]

1. What (type of) data will be collected or generated?
2. How will the data be described and documented?
3. Which ethical, legal and privacy-related regulations apply to this research?
4. Where will the data be stored?
5. What data will be archived after the end of the project and where?
6. Who will be responsible for the security and back-ups of the data?
7. Will the data be shared, and if so, with whom?

To prevent unnecessary bureaucracy, there is no policy of the FRT to centrally collect these plans. However, researchers are encouraged to compose a DMP for each of their projects – by way of personal diligence and promotion of scientific integrity. If the answer of the third question points to the presence of ethical, legal, or privacy-related concerns that may arise from a project, the researcher is responsible to submit the proposed project to the ethical commission of the FRT VU (*Commissie Wetenschapsbeoefening en Ethiek (CWBE) FRT VU*), in order to obtain advice relating to the design of the research. In case of any uncertainties, it is advisable to check with one of the faculty's Privacy Champions before submission to the ethical commission. VU-net also offers resources for composing a DMP, including concrete guidelines, an informative toolbox (open to everyone), as well as a tool for composing DMPs (only available for faculty staff).

[1] According to Tom Willaert, Dirk Speelman, Fred Tryuen, *Digitale geletterdheid: Dataverwerking in de geesteswetenschappen* (Leuven: Universitaire Pers Leuven, 2018), 25: '*data zijn representaties van observaties, objecten of andere entiteiten die worden gebruikt als bewijs voor fenomenen binnen wetenschappelijk onderzoek*.'

[2] Willaert et al., *Digitale geletterdheid*,26.

[3] Willaert et al., *Idem*, 41.

Data life cycle

Data have their own life cycle, usually characterized by several common stages. This life cycle tends to include: *planning* data collection, *collecting* the data, *processing* the data, *archiving* the data, and *sharing* the data. These stages will be discussed below, but more information is always available on the website of the University Library, see also *Useful DMP-related information*.

Before data is collected, the starting point should be: what data play a role in my research? And, besides this question, particularly the following two questions related to the DMP:

- How are these data described and documented?
- Are there certain ethical, legal and privacy-related requirements this project needs to meet? With empirical research this will usually be the case, while for projects of a more literary, historical or systematic nature this generally less so. If the answer to this question is positive, the ethics commission should be contacted (the faculty's Privacy Champions are also available for initial advice).

In principle this phase (of planning) is the point at which a DMP is composed (see above).

Following the planning-stage, in whatever (sub)discipline of the FRT the research is conducted, there is always a stage of data collection. A distinction can be made between raw, processed and analysed research data. All supporting data in the research, such as research logs, codebooks and metadata is called research documentation, which is the fourth kind of data. It is particularly relevant to distinguish between raw data that will be incorporated into the final research output and data that won't be. The latter type, material that will not be in the final publication, is still data; this concerns data that purposefully remains unused (for the final output). This material is equally relevant for the research, because a choice was made not to use said material, and the researcher should be able to justify this choice.

In the following stage, the data is processed. Even the slightest changes in layout or cleaning of the data turns raw data into processed data. Processed data is always regarded as a separate data asset from the raw data it comes from. This is the stage wherein the actual research is conducted. This can vary from the analysis of textual data to, e.g., processing sensitive, prosopographic material. For the question how this stage is to be realised (for a particular project), one is referred to the specific, methodological rules, operant in each (sub)discipline.

Processed data can be turned into analysed data by transforming the processed data into another representation of the data. A common example is turning an excel sheet into a graph. The graph is the analysed data, in this case. Other examples are syntaxes, code, or graphic representations.

Particular care should be taken when storing and sharing data (see more on this below).

Data storage and archiving

During the life cycle of data, safe storage is of particular importance. The University Library offers guidelines for safe storage. Researchers that are employed by the VU, or have a guest-status, are able to store their data safely on university hard drives. For sharing data within faculty research groups, a shared project-folder can be created on the G-disk (G:/). Safely dispatching data is possible through SurfFilesender, or – for uploading your data to the VU remotely – through Surfdrive or Edugroepen. the VU does not allow the usage of free cloud-applications (such as those offered by Dropbox, Amazon, or Google) for files related to VU-research; it cannot be sufficiently guaranteed that these providers satisfy the demands of EU privacy legislation, or that the VU can effectively exercise ownership-rights over files stored on these applications. Cloud-storage is facilitated, by the VU, through Surfdrive and Google Suite for Education. The latter is easily accessed by logging into Google with a VU-email. Sensitive (personal) data should only be stored on Surfdrive (NOT Google). The security risks of loose USB-drives should be prevented by storing and sharing files through the digital environment as much as possible.

In terms of archiving data (after the end of a research project): in principle it is advisable to archive data for at least 10 years, unless specific legal or discipline-related guidelines specify other terms (e.g. with personal data). If research was published (in which certain data was used), this data needs to be archived within 3 months of the official publication date. The costs of archiving are to be discussed within the faculty; the VU does not fully cover the costs for datasets larger than 50GB, hence this should be considered in potential subsidy applications. The University Library offers an overview of the options for mid- to long-term data storage.

PURE offers good options for creating a data repository that is also shareable (through a Digital Object Identifier or DOI). Larger datasets can be stored at DataverseNL or – in case of sensitive data – with the DANS-institute (Data Archiving and Networked Services). In terms of shareability of data (after a research project has ended).

For the shareability of the data (of a finished research project) – which is relevant for replicability – the VU uses the FAIR-principles: Findable, Accessible, Interoperable, and Reusable. These principles are also in agreement with those of Open Access.

Since students are largely excluded from the resources mentioned above, as much superfluous data should be deleted after the completion of research (e.g. for a thesis). Data that is traceable to individuals should be deleted. Anonymous (or anonymized) data should be stored safely – Surfdrive is available for students– and should be available up until graduation, after which it can be deleted as well (if the research has remained unpublished).

Personal data management

Special consideration should be given to the collection, processing and storage of personal data. Informed consent plays an important role when collecting personal data. Respondents or data subjects (e.g. when conducting interviews/using questionnaires) should be provided with all information, concerning the research project, that is required for them to make an informed decision on participating. They will need to explicitly consent to the use of the information they provide; to this end they need to fill out or digitally agree to an Information and Consent Form. Creating an Information and Consent Form should be done under the guidance of one of the faculty's Privacy Champions. The form should at least contain: the purpose of the research, the reason and method of collecting personal data, the intended use of this data, and who has access to the data. For students, a checklist with elements that should be included in your information and an example of a consent form can be found in the appendix.

In case of existing (secondary) personal data, one should make sure the organization providing said data is allowed to share it. When collecting personal data without the involvement of the person(s) one should weigh the privacy rights of the person(s) and one's own legitimate interest, and make a balanced decision.

In case of participant observation among groups or in an organizational setting, permission of the leadership of the group or organization to conduct the research is needed. As consent with conversation partners in informal research settings is not always possible, one should carefully weigh various interest beforehand. As a rule of thumb, there is always the ethical obligation to avoid any harmful impact on the subjects of a study. At all times openness and clarity about one's position as a researcher is important. For students: teachers or thesis supervisors should be consulted when making complex decisions related to personal data collection (especially when consent is difficult to obtain). When in doubt, you should check with one of the faculty's Privacy Champions.

Necessity is an important benchmark for the collection of personal data. Hence, when working with personal data, the planning-stage of research is all the more relevant. Students should take particular care in this regard, and discuss the necessity of collecting certain data with their teacher(s)/supervisor(s). For example, different standards apply to audio or video recordings, respectively, (junior-)researchers should be vigilant while obtaining the correct permissions for various types of personal data.

In so far as and as long as personal data is stored on private devices (recording devices, PC, USB, etc.), measures towards securing access should be taken. The data should be pseudonymized or anonymized as far as possible. Encryption and/or protection with strong passwords should be used to safeguard the data.

For the analysis of personal data, the researcher should make sure to use (offline) programs that do not run the risk of sharing the data. Programs like SPSS, R, Stata, Atlas.ti can be used. For online surveys, Qualtrics is advised. Any data covered by article 9 of the GDPR[4] should receive additional care – when being stored and/or analyzed – encrypting these data (using 7Zip or BitLocker for storage, and using Zivver when sharing) is highly recommended.

Personal data (that has not been anonymized or pseudonymized) should NOT be included in any research output – including a thesis – unless, and only in so far as, the respondent has explicitly consented to that datum being shared publicly. Anonymized and pseudonymized data can be used – given it cannot be traced back to any natural person – as long as this was made clear to the respondents while obtaining their informed consent. For theses or papers, the use of personal data may be inevitable, in which case it should only be shared with the teacher(s)/supervisor(s). Students should note that MA-theses will be published online by the University Library, if students want/need to opt out of this, they will have to discuss it early on with their supervisor(s).

Responsibility and ownership of data management

Every researcher is responsible for their own data management; this includes senior-researchers and PhD-students. In case of students (BA and MA), the principal responsibility is with the teacher(s)/supervisor(s). Hence it is relevant for teachers to be involved with the research of their students, particularly those aspects that concern data management. Students, however, are and should be encouraged to take their own responsibility with regard to the data they collect, process and store. If teachers want to use data collected by their students for further research, they will be responsible for the adequate storage of said data. In case of (meta)data connected to theses, the University Library facilitates the storage of said data (for up to 7 years).

In principle the party that finances a research project (either the VU or an external party) has the ownership rights of the data generated by said research project. The norm for researchers should be to strive for Open Access as much as possible (except where this is not feasible, as with personal data). In case of shared research projects, researchers leaving the VU, or PhDs separate case-by-case arrangements should be made for the (continued) management of the data.

External PhDs

The FRT works with a lot of external PhD-students, who usually do not use the internal VU IT-infrastructure. Hence, the composition of a DMP is a mandatory part of a PhD-proposal. Special care is required in these cases, especially when privacy sensitive data is concerned. This is a point of active attention for the Graduate School FRT.

Support

The Administration Office, IT, and the University Library of the VU offer a shared programme to support senior-researchers with RDM-related questions: the VU Research Data Support programme and RDM research support portal. Here you find online tools for writing a DMP online, for enabling end-to-end encryption in Surfdrive, and for sharing data with colleagues outside of the VU.

Embedding in the organisation of the FRT

Researchers working at the FRT are responsible for their own data management and DMPs, and should pay attention to the guidelines of this protocol. **For questions concerning privacy-related data,**

[4] 'All data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and all 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'.

researchers can turn to Rita van der Schriek-Hermans. The VU recommends every faculty to appoint a *data steward* and a *coordinator security*. At the FRT, these functions are fulfilled by the ethics and science commission (*Commissie Wetenschapsbeoefening en Ethiek*). The liaison for this commission is the vice-dean and dean of research: Katya Tolstoj.

Useful DMP-related information

- Video Scientific Integrity & Research Data.
- Video Research planning: Data Management Plans.
- Choose a template for the DMP that you will write (from the VU or a grant provider, see https://dmponline.vu.nl/).
- Section Data Management Plan on the LibGuide Research Data Management. the following sections help to relate important RDM aspects to the data life cycle:
  - Overview for the context.
  - Everything under Plan & Design provides information on all aspects required to initial phase of a project.
  - Everything under Collect & Store provides information on collecting and storing data.
  - Video Collecting, Recording & Managing data.
  - Everything under Selecting Data & Data Archiving, including the video Archiving & Publishing Research Data.
  - Video Data Management: Data Citation.
  - Everything under Publish & Share, including the video Persistent identifiers and data citation.
  - UGent Data Stewards. Knowledge clip: FAIR data principles provides information on the properties of data repositories that help making data FAIR.
  - Michener, W. (2015). Ten Simple Rules for Creating a Good Data Management Plan (9p). A practical guidelines for what a good DMP should contain.
  - Research Data Management VU Guidelines for information security 1.9.7.pdf (7p), sections 3.2 and 3.3. focus on the storage, sharing and archiving facilities.
  - VU RDM policy (RDM-policy-VU-2020-EN-v2.0.pdf, 4p).
  - Wilkinson, M.D. et al. (2016). The FAIR guiding principles for scientific data management and stewardship (9p). Information on where the FAIR data principles come from and how the various principles aim to achieve the goals of FAIR data management. Note that it's theoretical and doesn't provide practical advice on how to make data FAIR.

Annex I – Checklist and example of consent form

As mentioned previously, when you conduct interviews or use questionnaires, you should provide information about your research, and you need the permission (consent) of the respondents for using the information they provide.

In the information and consent form you:

- explain the purpose of your research;
- explain why and how you collect personal data. Also, explain that these data will only be used for your research and explain who has access to the data;
- explain that the anonymous/coded data is stored until you graduate;
- explain that the respondent can always withdraw his/her consent;
- are transparent about any risks that might be part of your data collection.

The information and consent form contains the following required elements:

- The title of your study.
- Confirmation that the information and consent form is read.
- Confirmation that there was room for questions by the respondents.
- Reminder on the voluntariness of participation. The right to decline to participate and withdraw from the research once participation has begun, without any negative consequences, and without providing any explanation.
- Permission for participation in the study. This permission must be voluntary and unambiguous.
- Date, name, and signature of the participant and researcher

The information and consent form is typically used for respondents who are at least 17 years old and mentally competent. When you work with people under the age of 17 or mentally incompetent adults, contact your supervisor(s) or teacher(s) for advice.

Example consent form

VU VRIJE UNIVERSITEIT AMSTERDAM

Student-researcher:

Supervisor(s):

Research:

1. I confirm that I have received and understood information about the study above and that I have been given the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am allowed to withdraw myself at any time, without giving a reason.

3. I confirm that I know that the information I provide will be anonymized. No personally identifiable information will be reported in any research product.

4. I agree to participate in this research project.

5. I agree that my interview will be recorded. Transcribed segments from the audio recordings may be used in published forms (e.g., journal articles and book chapters). In the case of publication, pseudonyms will be used. The audio recordings, forms, and other documents created or collected as part of this study will be stored in a secure location on the student-researchers' password-protected computers and will be destroyed after graduation.

Participant name:

Date of birth:

Interview date:

Signature participant: ___________________________________

Signature student-researcher: ___________________________________