

2024 Research Data Policy and Protocol, Faculty of Social and Behavioural Sciences, Leiden University

Faculty: Faculty of Social and Behavioural Sciences (FSW)

Version number: 11.0

Date: 2024-06-10

Status of this protocol: Approved by Research Directors, approved by Faculty Board

Written by: Céline Richard, Andrew S. Hoffman, Katie Hudson, Verena Ly, Willemijn Plomp, Jaap-Willem Mink

Privacy section in collaboration with: Max van Arnhem, Armin Hallilovic

Detailed review by (in alphabetical order): Peter Bos, Lotte van Dillen, Raymond van Erkel, Mitch van Geel, Cristina Grasseni, Wolfgang Kaltenbrunner, Thed van Leeuwen, Tom Louwerse, Annemarie Samuels, Marja Spierenburg, Wouter Veenedaal, Ludo Waltman.

Second round of review by (in alphabetical order): Doreen Arnoldus, Erik Bähre, John Boy, Roos van der Haer, Henriët Middendorp, Vincent Traag.

Approval process overseen by: Hester Bergsma.

Contact person for questions about this protocol: Céline Richard
(c.j.e.richard@fsw.leidenuniv.nl)

Location (Zenodo DOI): 10.5281/zenodo.12654228

Table of Contents

Table of Contents	2
Preamble to the data protocol.....	4
1. Introduction.....	4
2. Institutional Research Data Management Policies	5
3. Relevant legislation, agreements	6
4. Privacy requirements and research ethics.....	7
5. To whom this data protocol applies.....	13
6. Research data covered by the data protocol	13
7. Non-digital research data	14
8. Research software	15
9. Collaboration with third parties	16
Governance of Research Data Management in Social and Behavioural Sciences	16
10. Data Management Plans (DMPs).....	17
11. Publication Packages	20
12. Registering and managing research data not underlying a publication	24
13. Preserving non-digital data	25
14. Preservation and Deletion Policy and Protocol after the Initial minimal Retention Period	25
15. Exit procedure.....	25
Research Data Management practicalities	26
16. Guidelines for data collection.....	26
17. Guidelines for data transport and transfer	26
18. Guidelines for data storage	26
19. Guidelines for sustainable file storage	27
20. Guidelines for publishing research data.....	28
21. Guidelines on how to apply the FAIR principles.....	30
22. Integrating ethics review, privacy and Research Data Management in 7 steps for researchers	32
Responsibilities.....	34
23. Faculty Board	34
24. Institute Research Director	34
25. Supervisors and managers.....	34
26. Principal Investigator (PI).....	35
27. The researcher	35
28. Support within the Faculty	35
29. Central Support.....	37

30. Procedures for this data Policy and Protocol	37
Appendices	38
A. Glossary	38
B. Research Data Futures: Preservation and Deletion Policy and Protocol after the Initial Minimal Retention Period	40
a. Policy	40
b. Protocol	44
C. FSW research data storage guiding matrix	46
D. Links to Policies or equivalent documents from the Institutes of the Faculty referring to Research Data Management in whole or in part.....	47
1. Centre for Science and Technology Studies.....	47
2. Cultural Anthropology and Development Sociology.....	47
3. Education and Child Studies.....	47
4. Political Science	47
5. Psychology (please see 3. Education and Child Studies).....	47
E. Research Data Management infrastructure in development or that would benefit FSW researchers	48
1. Classification of existing infrastructure.....	48
2. New infrastructure	48
Bibliography.....	49

Preamble to the data protocol

1. Introduction

Amidst the proliferation of (inter)national legislation, University regulations, and funding agency policies calling for more responsible Research Data Management (RDM) practices within the scientific research system, researchers are now faced with ever-growing demands regarding the handling of their research materials – from safeguarding privacy, to devising methods to prevent data loss or corruption, to making them as findable, accessible, interoperable, reusable, and even as open as possible (but also ‘as closed as necessary,’ as the saying goes).

At the European level, the General Data Protection Regulation (GDPR) – and its Dutch instantiation, called the Algemene Verordening Gegevensbescherming (AVG) – specifies requirements for the handling of personally identifiable information, including in academic research settings. In the Dutch context, the Netherlands Code of Conduct for Research Integrity (KNAW et al., 2018) sets out a national-level framework for research integrity writ large, including a focus on Research Performing Organisations’ duties of care vis-a-vis good Research Data Management practices. Leiden University ratified its own Data Management Regulations in December 2021 (Regeling Datamanagement, henceforth referred to as RDM2021), which includes high level recommendations and requirements that apply to all employees, affiliates, and guests who conduct research under the sponsorship of the University. An important provision of RDM2021 is that each Faculty within the University develops its own Data Protocol, providing domain-specific elaborations of this latter policy.

In line with this provision, the present document is the first official version of the Research Data Policy and Protocol for the Faculty of Social and Behavioural Sciences (FSW) at Leiden University. This document is the result of a collaborative writing process involving Data Managers, Data Stewards, Policy Officers, and other Research Data Management experts from six of the University’s faculties. Working together, and off a common template furnished by the Leiden Digital Competency Center, we made great efforts to ensure alignment across the six Faculty Data Protocols while still accounting for the inevitable variations that arise based on the different ways that researchers approach their data – and thus their data management practices – across disciplinary and epistemological boundaries. In our own case, we have similarly done our best to account for the diversity of approaches to responsible Research Data Management that exists even within FSW itself, all the while remaining faithful to the spirit of the Leiden University Data Management Regulations. For this very reason, researchers may find the scope of this Policy and Protocol to be a bit broader than expected. However, such broadness is required to reflect the heterogeneity in the research carried across our five Institutes and extract commonalities. We direct researchers to their own Institute's Policies where they will find more actionable guidelines which are specific to their discipline or domain.

Given this situation, the Faculty Policy and Protocol is written in such a way that it speaks to Faculty, University, and national stakeholders in addition to Institute researchers themselves. It summarises the current commonalities in terms of Research Data Management across the five Institutes, accounting for the present landscape in terms of available infrastructure and

its limitations.¹ While this document should function along the principle of “apply or explain”, a final aim in fleshing out cross-Institutes commonalities is to reduce the need for researchers to “explain” by being equally inclusive of all of our research traditions.

As policies should be dynamic, and since the field of Research Data Management (RDM) is rapidly evolving, we expect that the present Policy will be updated regularly to account for changes in the wider ecosystem of RDM policies and infrastructures, as well as within FSW’s constituent disciplines themselves.

A list of general RDM definitions can be found in Appendix A. Several sections will describe discipline-specific requirements. Relevant definitions will then be given in the concerned paragraphs.

2. Institutional Research Data Management Policies

As previously mentioned, actionable guidelines which are specific to researchers’ discipline or domain will be found in the Institutes Policies (or equivalent documents). Every Institute of the Faculty of Social and Behavioural Sciences must have their own accepted Research Data Management Policy (or equivalent document) by the end of the year (December 2024).² The Institutional Research Data Management Policy (or equivalent document) should at least address the topics of Data Management Plans, a requirement from RDM 2021, and of Publication Packages, a requirement from the Deans of Social Sciences “Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences”. Please see the sections of this Research Data Policy and Protocol corresponding to Data Management Plans and Publication Packages for more details.

Each Institute must:

- Review the content of their Research Data Management Policy (or equivalent document) every two years with their Data Steward;
- Subsequently revise their Research Data Management Policy (or equivalent document) as needed.

The Research Director of each Institute is responsible for the review and revision of the Institutional Research Data Management Policy (or equivalent document).

Where relevant, the Institutes Research Data Management policies (or equivalent documents) are referred to within the main text and are linked in the appendices of the protocol.

Separate appendices have been added for research methods used across several Institutes that require specific Research Data Management (e.g., medical research as defined by the

¹ Appendix G presents the infrastructure the University is presently developing and the infrastructure the Faculty would like to see developed.

² This Policy and Protocol document will not apply retroactively – cf article 30. The exact start date of the DMP and Publication Packages requirements will depend on the Institutes Policies (or equivalent documents) and their respective implementation plans.

WMO, data processed on the SHARK and ALICE High Performance Computing (HPC) clusters).

3. Relevant legislation, agreements

A detailed list of the relevant legislation, whether they apply to all scientific disciplines (elaborating on RDM2021, art. 2) or are specific to Social and Behavioural Sciences can be found below:

- Legislation:
 - WMO (Wet Medisch-wetenschappelijk Onderzoek met mensen) applies for medical research (see [here](#) or full text [here](#)). This legislation concerns some research projects in the Institutes of Psychology and Education and Child Studies. Please refer to specific associated procedures in section of the Appendices.
 - [GDPR](#) (General Data Protection Regulation or [AVG](#) in Dutch) applies to any research dealing with Personally Identifiable Data. Please refer to specific associated procedures in section .
 - [Wet op het Hoger onderwijs en Wetenschappelijk onderzoek](#) (WHW)
 - [Archiefwet](#) (en bewaartermijnen)
 - [Auteurswet](#)
- National Guidelines
 - [2014 Dutch Code of Conduct for Scientific Practice](#) (Nederlandse Gedragscode Wetenschapsbeoefening) from VSNU (currently Universiteiten van Nederland)
 - [2018 Netherlands Code of conduct for Research Integrity from NWO](#)
 - [2022 Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences in the Netherlands by the Committee of Scientific Integrity, Data Storage and Reproducibility from the Deans of Social Sciences in the Netherlands](#) (DSW). This document will be referred to as GuidedSW2022 hereafter.
 - [NFU \(Netherlands Federation of University medical centers\) guidelines for medical research](#)
 - [Nethics code of ethics for research in the social and behavioural sciences involving human participants](#)
- University guidelines
 - [2021 Research Data Management Regulations \(Regeling Datamanagement\) of Leiden University](#)
 - [Leiden University Policy on privacy and information security](#)
- Faculty-specific guidelines and protocols
 - Ethics committees guidelines ([Education and Child Studies](#), [Psychology](#), [Social Sciences](#))
 - Institute-specific protocols – please refer to appendix
- Other Guidelines
 - Medical guidelines: [ICH – GCP](#). ICH-GCP is the Guideline for Good Clinical Practice from the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)
 - FAIR Principles (Findable, Accessible, Interoperable and Reusable) [1]

- CARE principles for indigenous data governance (Collective benefit, Authority to control, Responsibility and Ethics) (<https://www.gida-global.org/care> or [Carroll, S.R., Herczog, E., Hudson, M. et al. Operationalising the CARE and FAIR Principles for Indigenous data futures. *Sci Data* 8, 108 \(2021\).](#))
- TRUST principles for digital repositories (Transparency, Responsibility, User focus, Sustainability and Technology) ([Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 \(2020\).](#))
- [Belmont Ethical Principles](#) (Respect for Persons, Beneficence, Justice), Belmont report, USA department of Health, Education and Welfare, 1979

4. Privacy requirements and research ethics

4.1. Privacy

Research privacy aspects should be considered within the framework of the General Data Protection Regulation (GDPR), Europe's primary law for data privacy and security. The GDPR applies when processing **personal data** of EU/EEA citizens (please note: the GDPR also applies when processing data outside the EEA if the processor is within the EU/EEA² and vice versa).³ "Personal data" refers to any data that could potentially be used to identify a living person (GDPR article 4).

"Processing" refers to a broad set of acts performed with the data, including (but not limited to) storage, deletion, writing or copying of data (GDPR article 4).

In short: If you are working for Leiden University and working with data connected to living and identifiable persons the GDPR will apply.

4.1.1 Legal responsibility

Final legal responsibility for complying with the GDPR falls on Leiden University as a whole. Researchers must comply with the GDPR as members of staff and part of the research institution of Leiden University. Please note: In some cases, the researcher or Leiden University may act as a data processor⁴ and not a data controller⁵—this can happen if the research is paid for by an external non-academic party.

4.1.2 Terms and explanations

Any information that relates to a living individual and could feasibly identify that person is personally identifiable data. This includes the obvious such as name, phone number and (email) address, but also student number, IP address, audio recordings, nationality, or date of birth.

³ For more information, please see <https://gdpr.eu/what-is-gdpr/>.

⁴ The data processor only processes personal data on behalf of a data controller.

⁵ The data controller determines the goal and means for which the data is processed, and may request another party (a data processor) to process that data for them.

- **“Potentially” identify:** Be mindful that data that may not identify an individual *on its own* (e.g. gender, location, current work) may be used to identify individuals *in combination* with other identifying (or non-identifying data) or if that data is considered unique. For example:
 - Shoe size will not normally identify a person; however, if the data includes a note that subject #317 has unusually large feet (e.g., size 50) and is from Amsterdam, identification will be more likely.
 - Names will not always immediately imply a person’s gender but may do so if the name is specifically associated with one gender only.
- **Inferred data:** Data that is algorithmically inferred may constitute personal data if the inference could directly or indirectly be linked to an individual. This is only true if the accuracy of the inference is high and depends on context.
For example: an AI-algorithm that predicts gender based on iris-scans when the original research purpose did not intend to collect this data is considered new personal data compared to the original dataset.
- **Exceptions:** The GDPR will not apply when processing data from the deceased, or if the data subjects (i.e. the research participants) would take an unreasonable amount of effort to identify. Please note the latter only holds true in unusual circumstances (e.g. isolated tribes who have no contact with the outside world).
Information that has consciously been made public or published by an individual will not be considered personal data for most intents and purposes (e.g. social media posts, news appearances, published material).
- **Secondary data:** Please see the section on “Secondary data usage” below.
- **Examples of commonly used personal data:**
The most commonly processed personal data for research purposes are:
 - Name;
 - Contact information (address, e-mail, phone nr, social media handles);
 - Age;
 - Nationality;
 - Video/Audio recordings;
 - Payment /bank details.

4.1.3 Sensitive data: Special category data (GDPR - article 9)

The GDPR considers some personal data to be a higher risk or impact to an individual if processed. Special category data are:

- Health data (including mental health);
- Racial/ethnic origin;
- Political opinion/affiliation;
- Religious and philosophical beliefs;

- Trade union membership;
- Genetic data;
- Biometric data;
- Sex life or sexual orientation;
- Criminal history or record.

Special category data may only be processed when additional requirements are fulfilled. A concrete and balanced argument must be presented explaining why the processing of special category data is necessary for the intended goal, and how this outweighs the subject's privacy rights. Additionally, further organisational and/or security measures must be taken to ensure the data is secured.

Examples of additional measures are access management, encryption, anonymisation or pseudonymisation.

4.1.4 Legal grounds: consent (GDPR articles 6-7)

When conducting research that involves personal data, a legal basis is required for processing this data. Consent is only one such basis; however, agency and transparency are main guiding principles in the GDPR, which is why Leiden University recommends using consent as the primary legal basis when processing personal data for research.⁶

For consent to be considered legitimate within the GDPR framework, it must meet several criteria. These criteria are that consent must be:

- freely given, specific, unambiguous and informed;
- provided in a transparent, understandable and legible manner;
- recorded in a demonstrable way.

Clarification of consent criteria

The above in general terms means that consent must be asked in a clear and relatively brief way, in a language the participant understands and that includes what personal data will be used and for what purpose.

Participants must be able to give their consent in an active way (e.g. no pre-ticked boxes, or "silence equals consent").

Asking for consent is a process and can be an essential part of building trust or rapport with participants. The requirement for recording consent is relatively flexible and allows for filled-out consent forms, email responses, or an audio/video recording of a participant consenting. If you require support or assistance in finding a suitable way to ask consent that does not hamper your research, please contact the privacy officer.

4.1.5 Legal grounds other than consent (GDPR article 6)

The GDPR offers alternative legal grounds and exceptions for research where consent cannot be asked without making the research difficult or impossible.

The most often used legal grounds in Leiden University research projects where consent is invalid, impossible or unreasonable are:

⁶ It is important to consider as well that consent is not just a (potential) legal requirement, but also often ethically warranted.

- **Public interest:** If a data controller (oftentimes a government body or public institution) requires personal data to exercise the official authority vested in them they are processing personal data in the public interest. This allows for the waiving of the GDPR-compliant consent requirement, but crucially requires the controller to still provide transparency to a data subject and will often require asking consent for ethical purposes. The ethics committee can help with respect to consent requested for ethical purposes.

If research cannot be conducted by asking consent as described by the GDPR, please contact the privacy officer, there is probably sufficient ground to consider public interest.

- **Legitimate interest:** Some data processing may be expected for the daily functioning of an institution, and therefore falls under Legitimate interests. Advertising the University to prospective students is one such example.

Guidelines for other legal grounds: Legal grounds other than consent require an overriding and specific argument why consent is otherwise impossible to ask. The following will be assessed to find out if a legal ground other than consent applies:

1. **Necessity fits purpose:** Why do you need to process a participant's personal data for your research, and is the personal data strictly necessary to do the research?
2. **Interest:** Is there a legal or legitimate interest on the part of the University or researcher to process the data?
3. **Participant rights:** One cannot override participant's fundamental rights and freedoms under the GDPR. How are these safeguarded or otherwise protected if consent is not asked?

Please contact the privacy officer for advice if your research would be impossible when asking participant consent.⁷ The Faculty acknowledges that other legal grounds may be necessary for some research projects.

4.1.6 Pseudonymisation and Anonymisation

A common way to mitigate the usage of personal data is to make it more difficult or impossible to link the personal data used to a participant. This is done through pseudonymisation or anonymisation methods.

Please note that anonymisation and pseudonymisation are different terms under the GDPR, and definitions may differ from their usage in other fields.

- **Pseudonymisation (GDPR article 4):** Pseudonymisation refers to any strategy used to make it harder to identify a participant based on a set of personal data. (Partially) replacing personal data with tokens, ID's, and blurring or masking data are examples

⁷ Discussions at the National- (including within the Nethics, Nationaal Ethiek Overleg Sociale en Gedragswetenschappen, <https://nethics.nl/>) and University-level over the other application grounds of GDPR, especially the public interest ground, are ongoing with the aim of broadening its usage within Universities.

of pseudonymisation. Pseudonymisation is more common in qualitative research or research that requires personal data intact for its purpose.

- **Anonymisation (GDPR article 11):** Anonymisation refers to altering the personal data in your dataset in such a way that it becomes impossible to identify the participant, either directly or indirectly, including by the researcher. This will likely mean the deletion of **all** identifying personal data from your dataset. Anonymisation is a rigorous process and will not be possible for some types of research.

4.1.7 Underage participants (GDPR article 8)

Any participants under the age of 16 are considered underage. A separate distinction is made for those under the age of 14. If your research participants are under the age of 14, consent cannot be granted by the participant and must instead be provided by the participants' parent(s) or legal guardian. If your participants are under the age of 16 but not younger than 14, participants can grant consent only if you also gather consent from a parent or legal guardian.

If you require templates or further advice on having underage participants, contact the privacy officer.

4.1.8 Secondary data usage

Research may use data gathered from existing sources instead of first-hand data. Such data, when combined, may reveal more about individuals than originally intended.

In such cases you may need to re-evaluate the privacy aspects of the combined research data. Please consider the following steps:

- **Original Consent:** Determine whether the original consent obtained for each dataset covers the current intended use for your research. Consent is covered in-depth here-above.
- **Personal profile:** Assess whether combining datasets could create a more detailed personal profile. This could constitute a new processing activity under the GDPR, and the data should be treated as “new” personal data.
- **Original Legal Grounds:** Ensure that the combination of data is still permissible under the original legal grounds.

If your combined research data could lead to a new processing activity, or if the consent/legal grounds no longer pertain to this new dataset, please contact your privacy officer (privacy@fsw.leidenuniv.nl).

4.1.9 Practical advice for processing personal data in your research

- **Data minimisation (GDPR article 5):** only collect the minimum amount of personal data required for your research;
- **Consent (LU privacy policy):** If you are able to ask consent for your research purposes, please use the template provided by the University or privacy officer as a basis.

Note that consent for some parts of the processing may be withdrawn by a participant until the last stages of your research.

If your research is not possible when asking participant's consent, please contact the privacy officer for advice on other legal grounds.

- **DPIA (LU privacy policy):** fill out a Data Protection Impact Assessment (DPIA) form. If you are unsure or need guidance on how to fill out a DPIA, please contact the privacy officer to help in the process.
- **Non-University storage (GDPR chapter 4, section 1 and chapter V):** choose your storage and analysis platforms carefully when storing or processing personal data. If you are using external (non-University) platforms you may need to take additional steps in order to ensure GDPR compliance (personal data should not leave the geographical application zone of the GDPR as other potentially conflicting privacy legislations may then apply).
- **Data sharing (GDPR article 28):** when sharing or transferring personal data to external parties, a data processing or sharing agreement will be required. Please contact the privacy officer for more information.
- **Data lifecycle end (GDPR articles 5 and 89):** Unless otherwise consented by the participants or duly justified impossible in your DPIA, pseudonymise / anonymise personal data at the end of your research. Please contact the privacy officer or the data steward of your Institute for more details.

Remark: Psychology and Education and Child Studies have specific policies on the destruction of conversion keys.

4.2. Ethics committees: definition and application

The Faculty has three ethics committees that can be consulted on / review research proposals on criteria concerning ethically responsible scientific conduct:

1. CEP: Psychology Research Ethics Committee – contact ethiekpsychologie@fsw.leidenuniv.nl. The procedure to apply for the Psychology Research Ethics Committee is described online, please see reference [2]. Please note that the procedure is regularly updated.
2. Ethics Review Board of the Institute of Education and Child Studies (ECPW) [3] – contact ethiekie.ipw@fsw.leidenuniv.nl;
3. CEM: Ethics Review Committee Social Sciences. Review committee for the Institutes of Cultural Anthropology and Development Sociology, Political Science and the Centre for Science and Technology Studies [4] – contact ethiekmaatschappijwetenschappen@fsw.leidenuniv.nl.

4.3. Medical research

The *Medical research involving human subjects' law* WMO (Dutch: Wet medisch-wetenschappelijk onderzoek) is applicable to research that meets the two following criteria:

- The project is medical scientific research (medisch wetenschappelijk onderzoek), where medical scientific research as defined by the CCMO (Centrale Commissie Mensgebonden Onderzoek) is “Medical/scientific research which is carried out with the aim of finding answers to a question in the field of illness and health (etiology,

pathogenesis, signs/symptoms, diagnosis, prevention, outcome or treatment of illness), by systematically collecting and analysing data. The research is carried out with the intention of contributing to medical knowledge which can also be applied to populations outside of the direct research population”.

and

- Participants are subject to procedures or are required to follow rules of behaviour.

Such research must additionally be submitted to the METC-LDD (Medisch-Ethische Toetsingscommissie Leiden, Den Haag, Delft) of the LUMC (Leiden University Medical Center) or another Medical Ethics Committee. Please see appendix for more details.

If you doubt whether your project is subject to the WMO or not, contact your ethics committee and / or the Faculty monitor.⁸

5. To whom this data protocol applies

RDM2021 applies to all employees and persons performing research under the sponsorship of Leiden University (art. 3). This includes external PhD candidates and contract PhD candidates, visiting researchers, retired colleagues and any other guests or partners who carry out research at the Faculty of Social and Behavioural Sciences.

Research conducted by bachelor's and (research) master's students falls under the formal responsibility of their supervisors.

6. Research data covered by the data protocol

The Faculty adopts the conceptualisation of research data from Sabina Leonelli, philosopher and historian of science. She defines

“ ‘data’ as a relational category applied to research outputs that are taken, at specific moments of inquiry, to provide evidence for knowledge claims of interest to the researchers involved. Data thus consist of a specific way of expressing and presenting information, which is produced and/or incorporated in research practices so as to be available as a source of evidence, and whose behavior and scientific significance depend on the context in which it is used. In this view, data do not have truth-value in and of themselves, nor can they be seen as straightforward representations of given phenomena. Rather, data are essentially fungible objects, which are defined by their portability and their prospective usefulness as evidence.” [5]

Practical examples of data targeted by this protocol can be but are not limited to:

- Data collected during experiments, data reused, data gathered from external sources, data obtained while processing and analysing data, intermediate data products, derived data products, etc.;
- Software, models, scripts and code developed for the purpose of conducting research or as a research output (self-developed, community-developed, licensed, etc.);

⁸ The Faculty monitor being presently the Faculty research data manager.

- Description of the experiment design;
- Descriptions of equipment, research set-ups if relevant;
- All metadata that are necessary to understand the research output (by humans and machines);
- Other relevant supporting materials.

GuideDSW2022 provides examples of data for researchers in Social and Behavioural Sciences:

- “Registrations derived from experimental research;
- Survey data from questionnaires completed within the framework of research (including longitudinal research), collected by the researcher themselves or by an external fieldwork organization;
- (Transcripts of) video or audio material collected within the framework of qualitative research (open interviews, observations)”

In contradiction to GuideDSW2022, we choose to exclude field notes / personal notes from this version of the Faculty research data protocol due to their debated status.⁹ Field notes can be a heteroclite mix that encompasses diary-like personal reflections of a researcher placing them in a grey area, potentially closer to a researcher’s writing than to research data. We trust FSW researchers to extract relevant elements from their field notes and include them in their research data or publications.

Personally identifiable data must be dealt with in accordance with the GDPR. Medical data must be dealt with in accordance with the WMO. Specific steps must then be taken, please see section and appendix respectively.

Using the framework defined in this paragraph, the Faculty leaves the exact operationalisation of the research data definition to the Institutes, so that the final list of included data is pertinent and actionable for their research domain.

Research data can be digital or non-digital (see article for specific regulations linked to non-digital data).

7. Non-digital research data

We encourage FSW researchers to:

- Digitize their non-digital data when possible and feasible;
- Use the shared physical archive of their Institute - if existing - to store their non-digitizable research data as much as possible.

Non-digital data concerned by the Faculty research data protocol include:

- Paper questionnaires/surveys and related forms – some not being digitized or digitizable;

⁹ GuideDSW2022 which presents itself as a guideline “which can be further fleshed out under the motto ‘apply or explain’”. When the Faculty chooses to deviate from the document, an explanation is thus provided in this Policy and Protocol.

- Objects, documents and artefacts collected during field work or laboratory experiments deemed relevant to keep as a research element according to the researcher;

We choose to exclude from this present version:

- Biological samples stored at FSW (collected blood, syringes of IMP (Investigational Medical Product) after administration, etc.). This category of data will be the subject of a specific appendix in future versions of the Faculty data protocol with specific guidelines for their destruction. Please contact your data steward for advice.
- Paper version of field notes/ personal notes of researchers for the same reasons as for digital field notes / personal notes (cf. paragraph);
- Policy documents and other grey literature only available printed (often used as literature sources);
- Paper versions of research project management documents. Such documents are subject to specific regulations in the case of medical research (please refer to WMO requirements or appendix);
- Hardware / setups used in laboratory experiments as their management already follows specific workflows for most experiments.¹⁰

8. Research software

As mentioned in section 6, software developed for research or as a research output is considered as part of the data produced by a research project.

At a national level, the e-science center and NWO have released in September 2022 the first version of a Software Management Plan (SMP) and related guidelines. The Center of Digital Scholarship is presently in the process of developing a Leiden University SMP template and we expect to have University guidelines as well.

In the meantime, at the Faculty level, we differentiate two kinds of software, “processual software” and “sophisticated research software”. “Processual software” is defined as code used to get things done while “sophisticated software” stands out as an academic output in itself due to its high level of refinement and value.¹¹

For sophisticated software, the present protocol asks from FSW researchers to:

- Version their software. Research Software is versioned, and ideally tracked in an explicit system for version control, e.g. git, to support software integrity;
- Choose a software-specific usage license (for more details, please refer to [6]);
- Document their software with relevant metadata that encompass at minima: the creators of the software, the name of the software, the date the software was "published" (the date of the release of the version), the identifier (if no persistent identifier is linked to the software, the URL – for example for software on GitHub), the version;

¹⁰ We refer to, for example, physical stimuli used in experiments (toys, cards, puppets, etc), simulators or gaming material (mainly VR simulators for VR experiments), robotic / electronic setup developed by SOLO for experiments.

¹¹ Definition from Thed van Leeuwen developed in the context of a discussion over research software at CWTS with Ludo Waltman, Dan Rudmann and Andrew Hoffman.

- Make their software citable whether it is Open Source or not. Ideally an example of the citation format can be added to the software metadata or in a separate file dedicated to citation. A suggested format is for example *“Developer, A. A., Developer, B. B., & Developer, C. C. (yyyy). Title of the software: Subtitle [Computer software]. Archive Name. Retrieved Month dd, yyyy, or version date and version number from https://URL”*

We also encourage FSW researchers to refer, when relevant, to their sophisticated software within their publication (in a broad sense of the term publication: academic journal publications, pre-prints, blogs, etc.).

For processual software, we encourage researchers to evaluate on a case-by-case basis what can be done keeping in mind that efforts should be proportionate to the expected benefits for the researcher, for the research community or for society.

More information on software citation can be found in this reference publication from Katz et al. [7] and more information on the broader topic of research software on the website of the Research Software Community of Leiden [8].

9. Collaboration with third parties

In the case of collaboration with third parties (e.g., with other public, or private non-governmental entities), collaborators need to agree which party will be responsible for Research Data Management. This should include how research data will be collected (if data collection is involved), processed, accessed, used, and stored, as well clarify issues around intellectual property rights, such as copyrights and terms of use. A formal, legal agreement, is not necessary in all cases.¹² When personal data as defined by the GDPR will be shared, please contact first the FSW privacy officer and please also consider relevant ethical considerations (your ethics committee can provide advice). For other individual cases, help is available as first line through FSW Research desk or through the Institute’s data stewards. Further documentation is also available in the individual Institutes’ policies and on the [LURIS website](#).

Finally, some forms of research regard participants/informants as co-creators of the research methodology and/or published output; or involve communities that assert authority to control data from or about them (see e.g., the CARE principles of Indigenous data governance). In such cases, the researcher holds the research data ‘in trust’ in a broader sense than the pure legal requirements and needs to carefully negotiate rights and conditions around data access, publication, and long-term preservation, as well as acknowledgment of authorship. These are ethical issues that can be discussed with the relevant ethics committee.

Governance of Research Data Management in Social and Behavioural Sciences

¹² One example could be the case of a co-authored paper by researchers of different Universities working on publicly-available data.

The governance of Research Data Management in Social and Behavioural Sciences is organised around two core elements, Data Management Plans and Publication Packages.

Other paragraphs included in this section reflect practical problems experienced by some of the Institutes and aim at providing toolkits to solve the related Research Data Management bottlenecks. They do not constitute requirements at the Faculty level, but rather recommendations that can be included and elaborated on - if relevant - in the Institutes policies or equivalent documents.

10. Data Management Plans (DMPs)

10.1. Definition

DMPs are formal documents that “describe the data that is used and produced during the course of research activities, where the data will be archived, which licenses and constraints apply, and to whom credit should be given” [9]. A DMP should clearly describe all decisions and measures taken to guarantee responsible handling of research data and, if applicable, long-term availability of the research data.

10.2. Responsibilities linked to DMPs

10.2.1 Writing and updating the DMP

The **researcher** is responsible for writing and updating the DMP as the research project develops and evolves. The collaborators and, if relevant supervisors and other colleagues participating in the project, must be involved in the discussion leading to the chosen Research Data Management strategy that underlies the DMP writing process. The data stewards of the Institutes can provide discipline-specific advice and support during the process of writing the DMP.

10.2.2 Reviewing and approving the DMP

The responsibility for reviewing and approving will be elaborated per Institute within the framework outlined hereafter. The Faculty recommends that Institutes consider systematically reviewing DMPs by the data stewards. However, no formal approval will be required at the Faculty level.¹³ In case a funder requests a stamp of approval, data stewards of the Institute can provide it.

10.2.3 Archiving the DMP

For archiving, the **researcher** should send at least a copy of the finalised version of their DMP (at the publication or writing stage) to their Institute’s data steward. The **Institute’s data steward** is responsible in return for its archiving as outlined in this document (see below).

10.3. When to write a DMP?

¹³ For the Institutes of Psychology and Child and Educational Studies, the DMPs are approved by the data stewards on behalf of the Ethics Committee of Psychology (CEP) and of the data manager of the Child and Educational Studies Institute (currently Mitch van Geel).

According to RDM2021 the initial version of a DMP must be written before research starts. The present interpretation of the Faculty is as follows:

- For grant-funded research, a DMP should be written in accordance with the funder's requirements. For large consortia, research partners should define which of them will assume this responsibility.
- For other kinds of research, especially research based on the same methodology in the social sciences, the Faculty asks for an umbrella DMP per research area. The exact nature of a research area must be defined in the Institutes' Research Data Management elaborations depending on the nature of the research being carried out.¹⁴

Such umbrella DMPs will be updated when research methods change or evolve. Depending on the Institute, meetings with researchers or research groups will be organised to prepare the umbrella DMPs. We expect that this process will be spread over time, possibly over a year, due to the workload involved for researchers, research directors and data stewards while the DMP infrastructure is still in development.

Remark: The case of research consisting of reading and interpreting existing literature where no data is created that is separately stored (i.e. 'data' is included in the publication directly) can be decided upon per Institute in their Policy or equivalent document.

- For non-funded projects from Psychology, where all research projects are reviewed by the ethics committee of Psychology, a DMP will not be asked from researchers until the DMP questions are integrated in the ethics committee tool (CEP tool).¹⁵

The timeline for writing a DMP should be elaborated on at the Institute level.

10.4. Templates that can be used for a DMP

The researchers of the Faculty of Social and Behavioural Sciences are encouraged but are not required to use the University DMP template.¹⁶ Most funders (NWO, ERC and Marie Curie projects, ZonMw,...) will accept the Leiden template as long as it respects the categories of their own template. In case of doubt please contact your Institute's data steward.

In the case of larger collaborations, the researcher may have to work with a template from a project partner.¹⁷

Resources are available to researchers of the Faculty to help them fill in their DMP:

¹⁴ A research area can indeed, depending on the Institute, represent the project of a single researcher (personal umbrella DMP for non-funded research of a single researcher), a group of researchers using the same methods, a research cluster, focal area,...

¹⁵ Researchers already have to go through questions regarding ethics and privacy within the CEP tool, adding a DMP separate from the current ethics submission would be a too important bureaucratic burden. The integration of the DMP questions in the CEP is an ongoing project that we expect to have finished before the next revision of this document.

¹⁶ The Leiden University template is due to be updated.

¹⁷ On a practical note, a template with fewer questions is not necessarily easier to fill in than a more detailed one. The lighter template will provide less guidance, but the same level of detail can in some cases be expected by the reviewing committees.

- Annotated templates from the Institutes of the Faculty with discipline-specific comments / suggestions, please see appendix ;
- Tips & Tricks document from the Center of Digital Scholarship [10].

10.5. When to update the DMP?

As previously mentioned, the DMP is a living document that adapts and refines as the research project develops and evolves. Ad hoc revisions of the DMP should occur in the following cases:

- A significant deviation in the research methodology used;
- A significant method or process change such as a modification of the data storage solution;
- A structural change in the project outline such as a new PhD project starting or the need to add a further different study to the project.

These revisions should ideally be shared with the relevant data steward.¹⁸

10.6. Procedures for archiving and preserving DMPs

Article 15 of RDM 2021 requires that DMPs are stored centrally within the Faculty or the Institute. The researchers of the Faculty will be asked to review their DMP at the publication stage or writing stage to make sure that it reflects the reality of the project before handing in the final version (version of record) to the Institute's data steward that will be archived for the minimal retention period of the related dataset (usually 10 years except for medical research projects).¹⁹

Versions of DMPs will be stored by the Data Management team of FSW until a University-wide solution is developed.

The archiving of DMPs concerns:

- The above-mentioned version of record. This version needs to be kept for as long as the research data. The end of the conservation period of the version of record of the DMP will result in the destruction of all versions of the DMP related to the research project (the version of record and the initial version and / or any updates that could be preserved as well).

Except for sensitive DMPs - which access should be limited for security reasons²⁰ - the DMP storage folders will be accessible to:

- The Faculty data manager;
- The Faculty data stewards;
- The Faculty privacy officer;
- The research director of the Institute or their delegate (to allow for quick access especially in the case of an issue with the GDPR or research integrity). If the research

¹⁸ Researchers from the Behavioural Sciences Institutes who make (major) significant changes to their DMP, must send their revised version to the data stewards, so it can be checked and approved again.

¹⁹ The retention period defined by RDM 2021 is 10 years after the research. For medical projects, this period is longer and depending on their risk level.

²⁰ For sensitive DMPs, access will be decided on a case-by-case basis by the researcher and the data steward of their Institute (one can think of defence-related projects that are ongoing in several Institutes).

directors want to give broader access to the DMPs for example in the context of a visitation, the researchers should be informed.

If any other party needs access to a particular DMP, the approval of the concerned researcher will be a pre-requisite.

10.7. Sensitivity of the content of a DMP

Some DMPs can be openly shared provided that they do not contain sensitive information (personal data, sensitive information about the University digital infrastructure that could cause a security threat,...) – this can be discussed with the Institute’s data steward. DMPs should indeed be considered as any other research-related document: “as open as possible and as closed as necessary”.

10.8. Relation of DMP to other relevant procedures and documents

Please refer to section for details regarding the links between DMPs, DPIA and ethics committees’ review.

11. Publication Packages

11.1. Definition

Publication Packages consist of all information needed to assess the results presented in a publication according to the principle of ‘retroactive accountability’ (GuideDSW2022). Publication Packages will be preserved for a minimum of 10 years according to RDM2021, except for research under the WMO umbrella, where different retention requirements apply.

The exact definition of a Publication for the purpose of a Publication Package will be elaborated per Institute, as channels to share and disseminate research results vary considerably within the Faculty, e.g. preprints, articles, books, book chapters, conference papers, blog posts, etc. As a minimal common requirement, Institutes should include in the definition of the Publication at least original research articles (excluding review articles) published in peer-reviewed academic journals.²¹

11.2. National context and goals of FSW elaboration

Publication Packages are a specific requirement for Social and Behavioural Sciences as defined by GuideDSW2022 which presents itself as a guideline “which can be further fleshed out under the motto ‘apply or explain’ ”. This guideline presents an ongoing tension between the stated aim of the document, archiving of research in Social and Behavioural sciences, and the listed constitutive elements of the publication package that reflect for some part audit-related concerns. GuideDSW2022 also differentiates quantitative and qualitative research which is

²¹ The GuideDSW2022 guideline define publications are all research publications listed in the Faculty’s academic annual report. This list does not exist at Leiden, we thus deviate from the definition of GuideDSW2022

limiting for mixed methods projects. After consultation and discussion with the Institutes, FSW chooses to deviate in several ways from this text. The reasons behind these deviations will be explained in the following paragraphs.

As a first deviation, FSW chooses to rephrase the objective of archiving research as Publication Packages. Publication packages at the Faculty will aim at ensuring a transparent way of preserving research underlying a publication, transparency referring here to the definition of the Netherlands Code of Conduct for Research Integrity.²²

In addition, Institutes of the Faculty have further specifications on their objectives that reflect their own research traditions:

- Social sciences publication packages aim at ensuring a transparent way of preserving research underlying a publication while replicability of the study and reproduction of the results are often not possible nor relevant especially for qualitative and interpretivist approaches.
- Behavioural sciences publication packages aim at ensuring a transparent way of preserving research underlying a publication with the additional goals of replicability, reproducibility, sharing and reusability, striving to make Behavioural Sciences research more reliable and robust.²³

A longer-term objective of FSW is that the choice of infrastructure used to host publication packages (whether it is co-developed or pre-existing) will help make research carried out within the Faculty more discoverable for example through sharing relevant metadata.

11.3. Who is responsible for Publication Packages?

11.3.1 Creation of a draft Publication Package

GuideDSW2022 states that the first author of the publication is responsible for the creation of a publication package. The Faculty chooses to deviate as it seems more natural that the corresponding author would carry this responsibility.

- In the case of the corresponding author being a research master student or PhD student, the responsibility lies with the supervisor(s).
- When the corresponding author is not a researcher of our Faculty, if:
 - The corresponding author belongs to a Social Sciences Faculty of the Netherlands, then they will also be subject to the Publication Package requirement. They will be responsible for the creation of the Publication Package according to their own institution policy and protocol.

²² “Transparency means, among other things, ensuring that it is clear to others what data the research was based on, how the data were obtained, what and how results were achieved and what role was played by external stakeholders. If parts of the research or data are not to be made public, the researcher must provide a good account of why this is not possible. It must be evident, at least to peers, how the research was conducted and what the various phases of the research process were. At the very least, this means that the line of reasoning must be clear and that the steps in the research process must be verifiable.”

²³ Detailed Publication Packages guidelines for the Institutes of Psychology and Education and Child Studies have been created and are regularly updated by the data stewards (see Appendix).

- In any other case, deviating from GuideDSW2022, the Faculty expects the Institutes to develop their own guidelines.

At minima, clear agreements are expected to be made on a case-by-case basis with collaborators regarding the process of caring for information needed to assess the results of a publication and the potential creation of a Publication Package. The balance between the means needed (complexity of processes, administrative burden...) versus expected benefits should be central in these discussions.

The person responsible for the draft should submit it to their Institute's data steward.

11.3.2 Review of a Publication Package

The Institute data steward is responsible for reviewing the Publication Package before its long-term preservation.

11.4. Constitutive elements of the Publication Packages and final format

The Institutes must define the detailed list of constitutive elements of the Publication Packages for their own disciplines to reflect adopted methodologies as accurately as possible (please refer to Appendix for more details). These elaborations should be structured around the following core elements:

- The Publication or its persistent identifier, most often a DOI. (The Faculty indeed strongly encourages its researchers to cross-link publications and Publication Packages through their persistent identifiers. As a reminder, Leiden University researchers have the obligation to register their scholarly publications in [LUCRIS](#) and attach a pdf to make their scholarly output Open Access, see the [University Library website](#) for more details.);
- A Data Management Plan (DMP) or another, similar or equivalent document;
- Project metadata that encompass relevant research design elements insofar this is not described in the publication itself;
- When possible and relevant research data and their related metadata.

In contradiction with DSW2022, we will not ask for the inclusion of administrative documents aiming at justifying compliance to the NWO code of conduct to avoid creating a climate of surveillance within the Faculty.²⁴ These documents do not correspond to the goals that the Faculty sets for the Publication Packages and can otherwise be retrieved (from University administrative systems) in the context of the investigation of potential research misconduct or audit.

Once a publication package is finalised, it will be frozen as a version of record (read only) and can only be further modified through the creation of new versions.

²⁴ e.g. "Any hard evidence of the period of time spent in the field (e.g. flight reservations, train tickets, etc.)." will not be asked.

11.5. Including relevant data or not in Publication Packages

The Faculty trusts their Institutes to define selection rules for their discipline and their researchers to select research data to include or not in their Publication Packages. In case of doubt, help can be found with the Institute's data steward and general criteria that can be considered by researchers as a self-assessment are:

- Is it really research data? (e.g., conference abstracts or presentation slides about the data)
- Are the data already archived as part of another project, or with a publication?
- Are the data directly relevant to the publication(s)?
- Are the data unique (impossible to recreate)?
- Are the data valuable? E.g., in terms of transparency, re-use, quality, originality, size, possibility to combine them with other data, scale, costs of data production or innovative nature?
- Are there any obligations for long-term storage?
- Are the data subject to intellectual property rights?
- Are the data subject to privacy or ethical restrictions?
- If long-term preservation is chosen for the data, is the chosen infrastructure adapted to their characteristics (estimated cost, size, desired accessibility during preservation, etc.)

11.6. Access conditions to the Publication Packages?

Access conditions to the files within the publication packages will depend on the nature of the research: they should be “as open as possible, as closed as necessary” [10] balancing between ethics requirements such as transparency, open access, the privacy of research participants, and the interest of researchers.

In practice, each element of the package should be either: open, closed or restricted.

- If access is restricted, a transparent access request mechanism must be defined that includes the signature of a pre-established Data Sharing or Data Transfer Agreement.
- For open and restricted elements, usage conditions and licensing must be included.
- If relevant, embargo periods can be used – during an embargo period, only the description of the dataset is published, while the data themselves are closed.

Help can be found with the data stewards of the Institutes.

Deviating from GuideDSW2022, access to restricted or closed documents included in Publication Packages will be granted only in the context of the investigation of a potential research misconduct and following the approval of the Faculty Board.

11.7. Open access research described in its entirety by linked Persistent Identifiers (PID)

For some social science projects, all relevant research elements to be included in a publication package are open, have a Persistent Identifier and link to each other. There are also projects that do not involve any data. In such cases, the creation of a publication package can be

questioned.²⁵ We aim at providing a documented advice for such cases in the next version of this protocol in collaboration with concerned researchers and Institute's management.

11.8. Minimal common protocol for Publication Package creation and long-term preservation

Specific guidelines for researchers as well as the choice of a long-term storage solution are decided at the Institute level within the common framework outlined in this paragraph. The Institutes should feel free to experiment to find the best-suited solution for their research - data stewards are there to assist in this respect.²⁶

11.8.1 Submitting a draft Publication Package

Researchers are expected to submit a Publication Package draft to the Institute's data steward(s) within one month of their Publication going live.

11.8.2 Review and long-term preservation of the Publication Package

The Publication Package is preserved after being verified by the Institute's data steward – this process can take time depending on the complexity of the submitted package and on the number of Publication Packages submitted concomitantly. The Institutes strive to take the necessary steps to start the long-term preservation of the publication packages within **three months** after the submission of the draft to the research data stewards.²⁷

12. Registering and managing research data not underlying a publication

At this stage, the Faculty will not register data not underlying publication. We invite the researchers to select from their non-published data the ones that are worth long-term preservation. If in doubt, the researcher can use the following selection criteria:

- Is the reason for the data not being published an unexpected problem during data collection or analysis. If yes, is it useful to document it / keep a record of it for future research?
- Are the data unique?
- Does the researcher foresee possible reuse by third parties?
- Would it cost a considerable amount of time/money to reproduce the data and how does it relate to the time, effort and funds invested by the researcher to set up the long-term preservation?
- Are there any obligations to preserve the data long-term (e.g. due to the funder)?

²⁵ We would be reluctant to create publication packages as a word file listing PIDs. If the purpose of publication packages is for audit of the Faculty and conservation of the work being carried under its umbrella, we feel that a more constructive way of creating publication packages in such cases would be to use knowledge graphs linking the elements of a publication.

²⁶ It can be noted that the Institutes of Psychology and Education and Child Studies have decided for data archiving and data publication on DataverseNL. The cost of the use of DataverseNL is covered by the University for all datasets < 1TB.

²⁷ GuideDSW2022 states one month but this timeline is often too short in practice.

If the answer to any of these questions is “Yes”, the researcher is encouraged to deposit the data in a trusted repository. Data stewards can support this process.

13. Preserving non-digital data

The preservation of non-digital data is organised at the Institute level with Faculty support. Please refer to the Institutes’ Policy (or equivalent document) for details or contact the data stewards for support.

14. Preservation and Deletion Policy and Protocol after the Initial minimal Retention Period

Please refer to Appendix B.

15. Exit procedure

The Faculty recommends to its constituent Institutes to include within their exit procedures (“exitgespreks”) topics related to Research Data Management to avoid bottlenecks further down the line (such as potential need to delete the research data after the end of the minimal retention period, see article). A list of suggested topics to discuss that could be of interest for all Institutes is presented below:²⁸

- Does the researcher who is leaving need / want to carry on using the research data collected during their employment at Leiden University? If yes, the legal ground of custodianship for the concerned research data should be investigated and depending on cases, agreements must be made with the Institute to:
 - maintain access to the data for a defined period of time and / or transfer a copy of the dataset to the researcher;²⁹
 - have the researcher take the data with them.
- When (a copy of) the data is kept within the Faculty:
 - Where are the research data of the leaving researcher stored? Are they copied on a hard drive or are there paper versions? Irrespective of their formats, are the datasets documented?
 - Can the research data collected by the researcher be deleted after the end of the minimal retention period or are there arrangements in place for further preservation?
 - Who can access the dataset? Are there procedures in place for data sharing if relevant?
 - If the research data are pseudonymised, who retains access to the conversion list if it has not been destroyed? If the conversion list is still present, are there agreements regarding its destruction?

²⁸ A large part of the suggested topics is inspired by the “Formulier einde aanstelling”, section about research, from the Institute of Education and Child studies.

²⁹ The Research Data Management team is still investigating the range of possible legal frameworks depending on research types.

- If the results of the research obtained using research data collected at Leiden University will be shared and disseminated, what are the agreements made in terms of chosen media (preprint, article, book, blog, etc) and authorship?
- Is there any handover in relation to research data that has not been mentioned in the previous set of questions?

Research Data Management practicalities

This section presents general practical guidelines for daily work with research data. More detailed, discipline-specific guidelines will be found in the Institutes' elaborations.

16. Guidelines for data collection

As a general principle, during data collection, the storage time in less secure locations (e.g., recording device) is being minimised and personal data is being deleted or moved to a more secure location as soon as the research allows for this (e.g., phone records, e-mails with respondents/participants).

An important resource for the Faculty is the SOLO (Support for Research, Laboratories and Education) wiki [11]. It contains technical documentation and resources for the research technologies supported by SOLO, such as: software, hardware, physiology, facilities, participants payment procedures, VR (Virtual Reality) experiments, research on the High Performance Computing clusters as well as general support and safety guidelines. Data collection protocols for high-risk research (sensitive data, bio data, ...) are also available.

17. Guidelines for data transport and transfer

Researchers should transfer data as much as possible directly to their recommended/approved data storage solutions. Please consult the [University guidelines for sharing and sending files](#). The following additional guidelines are highly recommended:

- When transporting sensitive data on physical devices, encrypted devices should be used or other mitigating measures need to be taken.
- University-managed hardware should be used wherever possible.

When transferring sensitive data, the transfer protocol must use encryption. If personal data falling within the scope of the GDPR are concerned, the chosen transfer solution should not move the data / copy the data outside of the GDPR application area - another potentially conflicting legislation may apply.

For a comprehensive overview of things to consider when collecting data 'in the field', please consult Leiden University's [checklist for Data management while working abroad](#).

18. Guidelines for data storage

As mentioned previously, data storage refers to the storage of data during research. A guide summarising the preferred data storage options in function of the discipline-specific requirements of the researchers is presently being developed at the University level. In the meantime, as a temporary solution, FSW's information manager and data manager have

drafted a data storage decision matrix, cf appendix . Data stewards are always available for advice. To summarise, it is preferred that researchers:

- Store their data on J:/ drive research data if possible.³⁰ In a first instance, it will allow for the Faculty to make an inventory of the data needs of its researchers, to provide more tailored advice and find appropriate solutions.
- Avoid as much as possible devices that cannot be wiped out from a distance in case of loss. This creates security risks, especially if the researcher deals with sensitive data.
- Avoid Tools / solutions where data will be stored outside of the European Economic Area (EEA) as the data are then subjected to the local legislation which is not always compatible with the European / Dutch legislations.³¹ If possible, research data should ideally be stored within the Netherlands.
- For collaborative projects, avoid using personal accounts of University-provided resources such as P:/ drive, OneDrive or SURFdrive as it can create problems of access to the research data if one of the collaborators moves to another institution.

It has to be noted that P:/ drive is presently being phased out as storage for research data within the Faculty. We expect researchers to not store new research projects on P:/ drive during the transition period. This transition has started already for Education and Child Studies and Psychology and will start in the coming year for the social sciences Institutes.³²

If in doubt, the researchers can contact their data steward for advice.

19. Guidelines for sustainable file storage

- **Data formats** that are durable and that are uncompressed or lossless should be preferred if possible.³³ Regarding durability two different choices can be made: either open and non-proprietary formats or ubiquitous proprietary formats. The ubiquitous proprietary formats are indeed as likely to perdure as the open and non-proprietary ones if they are community standard. DANS curates a list of durable file formats: <https://dans.knaw.nl/en/file-formats/>.³⁴
- **Data documentation:** Stored datasets need to be well documented, as much as possible during the current phase of the research, according to the Publication Package requirements. This includes:
 - A brief description of the dataset (metadata properties);

³⁰ Other solutions are presently used by some researchers of the Faculty including Research Drive and YODA. Please contact your data steward for advice (for most solutions significant costs will arise when datasets exceed 1TB).

³¹ The EEA is the application scope of GDPR.

³² The transition from P:/ drive is organised by the Faculty Information Manager and the ICT coordinator.

³³ Uncompressed: the data are saved in their original form, without being compressed (e.g. for images: the TIFF format has an uncompressed option – the files become very large though).

Lossless: the data are saved in a compressed way to minimise file size but the original, uncompressed data can be recreated from the compressed version (e.g. for images: the JPEG2000 format is lossless while other JPEG formats have lossy compression; the PNG format and the compressed TIFF format are lossless).

³⁴ Examples of such interoperable formats are but do not resume to: .pdf for documented text; .csv for tabular data; .txt for text; .tiff for image files; ...

- An overview describing individual files, codebooks describing data elements used and descriptions of code syntax, where applicable.
- **Rich metadata:** Rich metadata contribute to the FAIRness of a research dataset (principles F2, I1 and R1 – see figure 1). They correspond to an accepted metadata standard in a machine-readable format (the most commonly used being Dublin Core). In practice, the creation of such a rich metadata file consists in filling in a simple a form that will then produce a metadata text (that can be saved as a readme file if necessary). This is called a rich metadata generator.
Some repositories propose to automatically create rich metadata when registering a dataset. If not, the researcher can use a pre-existing rich metadata generator. Data stewards can provide guidance and help in this process.
- **File naming, naming conventions and folder structure**
 - **File naming.** As outlined by TU Eindhoven (<https://www.tue.nl/universiteit/library/library-for-researchers-and-phds/research-data-management/rdm-themes/data-organisation>): Systematic file names help identify the correct datasets without having to open them. Good file names are:
 - Consistent (based on file naming conventions if any exist);
 - Distinctive (distinguishing between various file versions and files with similar subjects);
 - Indicative (meaningful).

The sequence of elements of which a file name is composed (e.g. subject/content description, date, version, file type, project number, research team) is relevant if you wish to sort your data files in any specific order. It is recommended to compile a README file explaining the meaning of the file names and keep it with your files and update it regularly.
 - Some disciplines have **naming conventions**. Please see the following reference for more details - mainly relevant for behavioural sciences Institutes [12]
 - A clear **folder structure** with usually a maximum of 3 to 4 layers of sub-folders is also important for quick and easy data file identification.
- **Revision control (ideally versioning scheme):** Proper version management should be performed on the data such that:
 - The original (raw, unmodified) version of the data can always be identified;
 - Errors when working with the data can be mitigated so that at the most one day of work is lost;
 - Versions of the dataset that constitute the basis for specific publications can be identified as versions of record.
 - for more details on versioning see [13]

20. Guidelines for publishing research data

Publishing research data refers here to making research data available either through publishing them in a repository, a public archive for a specific kind of data (e.g. National

archieff), a research data or software journal or through another route (e.g. YODA publication).³⁵

20.1.1 Persistent identifiers

When publishing research data, researchers should make sure that their data are granted a persistent identifier (for example a DOI) – this is most often the case when data are published in a repository or a public archive. A persistent identifier ensures that the data set stays findable and increases the chances of being referred to. Citing datasets participates in their discoverability.

Persistent Identifiers can also serve different purposes. According to University regulations, Leiden researchers are required to create an ORCID. The ORCID is an example of persistent identifier that can be added as a rich metadata element. Whereas a DOI is used to find and identify discrete research outputs such as a publication or a dataset, an ORCID is a unique series of numbers assigned to an individual researcher. The ORCID is particularly useful for (1) finding and identifying an individual researcher; (2) linking that individual to the contributions they have made to the scholarly record; as well as (3) disambiguating the identities of several authors who may share the exact same name.

20.1.2 Access conditions and licensing

Published research data should come with a clear and accessible access conditions and data usage license. For more details, you can contact your data steward or consult the following resource [14].

20.1.3 Choosing a repository

The choice of the repository should correspond to the common practices in the discipline and legal requirements.³⁶ In case of doubt, national repositories and trusted repositories (according to TRUST principles, ideally having a CoreTrustSeal certification <https://www.coretrustseal.org/about/>) should be preferred.

The repository should ensure that:

- the access conditions of the data are clearly displayed;
- if an embargo needs to be applied, only the description of the dataset is published;
- standardised exchange protocols are used so that metadata are publicly accessible and harvestable by machines.

Important remark: Additional legal requirements should be met for sensitive data as defined by GDPR, the most important one being that the data should be stored within the EU / EEA.

³⁵ Should researchers want to use a journal for their research data, the Faculty strongly recommends to choose only for free open access journals.

³⁶ We aim at providing a documented list of suggested repositories in the future years after surveying the researchers of the Faculty to gather their discipline-specific habits. Some of the protocols and policies of the Institutes presently mention examples. In case of doubt please contact your Institute's data steward.

When dealing with sensitive data the privacy officer should be consulted if any doubt regarding the choice of a repository arises.

21. Guidelines on how to apply the FAIR principles

21.1. Introduction to the FAIR principles and goal

The FAIR principles (Findable, Accessible, Interoperable and Reusable) [1] are a set of guidelines for data management and stewardship released in 2016 by a group of stakeholders representing academia, industry, funding agencies, and scholarly publishers. FAIR is originally not an acronym, but a set of principles (see figure 1) that aimed at filling the gap between the ambition of “data and knowledge integration and reuse by the community after the data publication process” and the poor existing digital infrastructure that surrounded scholarly data publication in 2016, when the principles were first published. FAIR principles are thus firstly and explicitly aimed at machines, or computational stakeholders, that support individuals. This entails important consequences in terms of underlying digital infrastructure adapted to data management.

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Figure 1: FAIR principles as described in the original 2016 article (Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. *The FAIR Guiding Principles for scientific data management and stewardship*. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>)

Since the publication of the principles, the digital ecosystem surrounding research data has evolved. It is, however, still a fast-evolving field both nationally and internationally.

We acknowledge in this protocol that on a University level and on a Faculty level not all of the necessary digital solutions and tools exist to fully implement the FAIR principles as they were initially intended. The recommendations presented hereafter should thus be considered as a

starting point that will evolve in future versions of this protocol as tools become available in a way that meets the needs of diverse research communities.

Open data is a requirement that cannot be met by most of the data dealt with and produced by researchers in social and behavioural sciences due to their sensitive nature (as detailed in section). However, FAIR does not equate to open. An important part of the data of the Faculty, though not all, can be made FAIR to some level.³⁷ For sensitive data, sharing rich metadata indeed ensures some level of findability, accessibility, interoperability and reusability. The following paragraphs will describe this process.

21.2. FAIR: findability of data

To ensure the findability of data, researchers are encouraged to deposit their data in a repository that in addition to above-mentioned characteristics:

- (1) grants a persistent identifier (for example a DOI) to the dataset;
- (2) that provides or allows for rich metadata corresponding to the dataset; and
- (3) that indexes this metadata such that it can be found via multiple search channels (e.g., within the repository's own catalogue, but also via general search engines on the web):

In some cases, project metadata hosted by a repository will be limited to whatever description of the data is provided by the submitting researcher(s) while in other cases repositories offer services to further enrich the existing metadata using automated tooling.

If depositing data in a repository is not possible, depositing metadata only could ensure the findability of the research.

21.3. FAIR: accessibility of data

The accessibility of data is a function of:

- (1) the ability of search tools to retrieve (meta)data using persistent identifiers (as outlined in §9.1); and
- (2) the existence of and adherence to procedures that outline what actors, and under which conditions, a given dataset can be accessed by a third party. Even in cases where a specific dataset has been deleted, destroyed, or is otherwise no longer available, the metadata for that dataset should remain accessible via the same channels, and under the same conditions, initially used to access the dataset itself.

In practice, the accessibility of the dataset can be ensured mainly by choosing a trustworthy repository (see 20.1.3).

21.4. FAIR: interoperability of data

To ensure interoperability of the data the researchers are encouraged to:

- Use rich metadata as described above;

³⁷ We acknowledge in this protocol that data stewards of the Faculty have already come across projects for which valid reasons exist preventing the researchers from implementing any type of sharing of their research data and more remarkably in some cases even of the metadata corresponding to the research data. Due to their sensitivity, Research Data Management for such projects should be discussed with the data stewards of the Institutes.

- Use data formats that offer the most interoperability (cf data formats)

21.5. FAIR: reusability of data

To ensure reusability of the data, the researchers can make sure that:

- Any data comes with a clear and accessible data usage licence (see above);
- Clear and well-documented naming conventions for the files within the dataset including versioning are used (see above);
- That the rich metadata meet research-domain relevant community standards if any exist.

22. Integrating ethics review, privacy and Research Data Management in 7 steps for researchers

The purpose of this paragraph is to help researchers save time by integrating the various steps they have to go through. These steps for the different requirements can overlap so that addressing them at once can save a substantial amount of time.

1. Inform your research design

Briefly review the available documentation about privacy, Research Data Management, and ethics. This will help speed up each respective review process.

- **Privacy requirements:** Personal data is collected for the majority of research projects carried at FSW. You will likely need to ask for consent from participants, and may need to pseudonymise or anonymise data. [More information can be found here.](#)
- **Research Data Management:** Review your Institute's (or section's if relevant) Research Data Management policy / protocol. [More info can be found here](#)
- **Ethics Committee:** Find out which Ethics Committee is responsible for reviewing your research. Ethics Committees' main task is to give advice to researchers and to review research proposals on criteria concerning ethically responsible scientific conduct within the Institutes. This encompasses but does not limit to cases where the privacy of research subjects is concerned, aspects of human dignity during interviewing, the acquisition of mutual consent and a possible conflict of interests in the case of third money flow i.e. contract research. [More information can be found here.](#)

Make information concrete: You should now have a basic idea of what information is potentially required by the privacy officer, research data steward and ethics committee. Make sure this information is discussed and agreed on by your team.

Please note: The following steps 2, 3 and 4 should be done in parallel as the Ethics Committee will give approval on aspects of ethics, privacy, and in some case on aspects of Research Data Management.

2. Privacy aspects

- **Organise Consent Forms:** If you need to ask for consent from participants, prepare clear and comprehensive consent forms. Review requirements for a consent form or use a template provided by your privacy officer.

- **Contact the Privacy Officer:** The Privacy Officer can help with any privacy-related queries or issues. You may need to conduct a Data Protection Impact Assessment for high-risk research.
- **Data Sharing/Data Processing Agreement:** If you are sharing data with an external (i.e. non-Leiden University party) you will likely need to establish an agreement with that party. If you are using third party data, you are likely to have to sign a data sharing/processing agreement. The privacy officer will help you go through this process.

3. Research Data Management

Check your Institute's Research Data Management Policy (or equivalent document) to know if you need to work on a DMP in relation to your ethics submission (it might not be the case). Contact your Institute's research data steward: the Data Stewards and Data Manager are available to assist you with any Research Data Management questions you may have and to help you find relevant Research Data Management solutions for your research.

4. Ethics Application

- Check your ethics committee submission process and fill out the ethics review form (/ use the CEP Tool)
- Include relevant Research Data Management and privacy documentation. Any privacy question must be submitted and answered by the Privacy Officer before final approval can be granted by the Ethics Committee.

5. Research phase

- **Carry out your research:** Gather and / or analyse data being reflexive about the process and maintaining the standards that you have set for your project during your research design journey.
- **Amendments:** If you need to deviate from the approved research design, check with the Ethics Committee if an amendment is required. Do not forget to update your privacy documentation as well as your DMP (which is aimed at being a living document).

6. Sharing results, publication and long-term preservation of research data and software

- **Prepare for publication and / or sharing your results**
- **Publishing:** Should you choose to publish your research, please check the Publication Packages policy of your Institute.
- **Sharing your results** can also be done in a less traditional fashion (blog post, etc). Should you still want to create a Publication Package, the research data steward of your Institute can help.
- **Long-term preservation of research data and software:** Do not forget to consider aspects such as documentation (metadata), versioning, data sharing if applicable (repositories, access conditions, licences,...), secure storage, minimal retention periods, anonymising data if applicable, etc. Contact your research data steward for support.

Responsibilities

23. Faculty Board

The Faculty board's responsibilities are to:

- Provide means and support for the elaboration, implementation, review and regular evaluation of the research data protocol;
- Review and organise the evaluation to be scheduled every two years or when revisions are needed to remain compliant with governing regulations, policies, ... ;
- Foster recognition of responsible Research Data Management;
- Outline expectations for permissions and deletion decisions on specific research data;
- Disseminate Research Data Management vision and outlook to all researchers;
- Encourage a culture of mutual / shared intellectual curiosity between researchers and Research Data Management staff for each other's work;
- Delegate their responsibilities or give mandate where relevant to the research portfolio holder (Vice-Dean for Research presently).

24. Institute Research Director

The research director is accountable to the research portfolio holder and is generally responsible for Research Data Management within their Institute.

The responsibilities of the research director are to:

- Foster the elaboration, regular update and adoption of the institutional research data policy and protocol;
- Make arrangements for the final responsibility for research data produced in the Institute, such as decisions about access, deletion of data etc. Such arrangements can be dependent on the development of University-wide and/or Faculty-wide solutions;
- Contribute to the elaboration of the Faculty research data protocol;
- Delegate their responsibilities or give mandate where relevant to staff members or researchers³⁸.

25. Supervisors and managers

The supervisors and managers are accountable to the Institute Research Director.

Their responsibilities are to:

- Develop, maintain, and disseminate Research Data Management procedures for their section or team (e.g., research managers, facility managers, collection managers, data analysis managers, etc.).
- As a practical step, to avoid orphaned data or data loss, the supervisors and managers should know where the research data of their team members are stored and if possible be able to access them. If not implemented during the course of the research project, the supervisors and managers should make sure that the transfer of knowledge and

³⁸ These responsibilities are effectively delegated for a large part to the research directors at FSW.

responsibility for the research data is done at the end of the project when project members leave the University for example during an “exit conversation” (exitgepsrek).

26. Principal Investigator (PI)

The Principal Investigators are accountable to the Institute Research Director.

Their responsibilities are to:

- Inform staff members and students about the general data management policy;
- Develop the DMP;
- Notify all relevant parties in case of changes in the research method and to update the DMP accordingly;
- Carry end responsibility about the management of the research data;
- Delegate responsibilities or gives mandate where relevant to researchers.

27. The researcher

The researcher, if different than the PI is accountable to the PI.

Their responsibilities are to:

- Manage research data according to the Faculty and Institution research data policies and protocols generally, and particularly as described in the DMP of their project.

28. Support within the Faculty

28.1. Support staff

- Data Management and Stewardship

The Faculty has a Research Data Management network that works collaboratively on projects related to Research Data Management as well as on trainings and workshops. The core Research Data Management network is composed of:

- ✓ The Faculty Research Data Manager and Policy Officer for Research Data Management (1.0 Faculty);
- ✓ The Research Data Steward of Cultural Anthropology and Development Sociology and of the Center for Science and Technology Studies / researcher on Research Data Management;
- ✓ The Research Data Steward of Political Science;
- ✓ The Research Data Stewards of Psychology and Education and Child Studies.

Contact is possible

- ✓ by email
 - datamanagement@fsw.leidenuniv.nl for general questions
 - Datastewards_PSY_PED@FSW.leidenuniv.nl for specific questions related to Psychology or Education and Child Studies
- ✓ Through Open Office hours:
 - General Open Office hours: first Tuesday of each month from 10:00 to 11:30 in 3A21

- Psychology and Education and Child Studies Open Office Hours:
Every Tuesday from 13:00 to 15:00 in 4A22

- Privacy officer – privacy@fsw.leidenuniv.nl

The privacy officer can help with questions about privacy, GDPR, and information management in the context of research, education, and operations.

- ICT coordinator and Security officer
- Information manager
- Data collection protocols for the Faculty laboratories – labsupport@fsw.leidenuniv.nl
- Policy officers

Psychology and Education and Child Studies Institutes have dedicated policy officers who can provide advice on research- and policy-related aspects.

- Ethics Committees

The Faculty has three ethics committees that advise researchers on ethical issues relating to their research, review and approve ethics submissions, see 4.2. It is to be noted that research falling under the umbrella of the Medical Research Involving Human Subjects Act (in Dutch: WMO) needs to be submitted to the METC-LDD of the LUMC or another Medical Ethics Committee.

- The Grant Support Office (Research Desk):
 - ✓ Provides information and advice on grant opportunities;
 - ✓ Supports researchers in the process of developing and writing their grant proposal and guide them through the mandatory administrative matters;
 - ✓ Can be contacted through: researchdesk@fsw.leidenuniv.nl

28.2. Training and information

- Research Data Management is now mentioned in the onboarding module of the Faculty for new researchers (including all students)
- Information can be found online:
 - Practical information on the [RDM wiki](#)
 - Other RDM info on the [RDM website](#)
- Discipline-specific documentation can be found in appendix
- Trainings and Workshops are available for the Faculty members on
 - Data management training course
Following such training is mandatory for PhD students of the Centre for Science and Technology Studies, Education and Child studies, Psychology. There are two different trainings:
 - ✓ The standard [CDS training](#)

- ✓ The discipline-specific training Psychology and Education and Child Studies. More information can be found on the Open Science webpage of [Psychology](#).
- Research Data Management and Privacy Session - a training given by the privacy officer in collaboration with the data stewards of Psychology and Education and Child Studies

29. Central Support

Central support is for the most second-line support, researchers are invited to contact Faculty support preferentially.

- The [Research Support Portal](#)
- The Center of Digital Scholarship from the University Library is the second line Research Data Management support - [UBL CDS](#)
- IT - [ISSC helpdesk](#)
- Security office
- LDCC (can be contacted through the data stewards)
- [Privacy Service Point](#)
- [Grant Development office \(formerly LURIS\)](#)
- [Knowledge Exchange Office LURIS](#)
- Documentary Information and Archive Management department [DIA](#) (Documentaire Informatievoorziening en Archiefbeheer)

30. Procedures for this data Policy and Protocol

This data Policy and Protocol takes effect as of the 10th of June 2024 and will not apply retroactively. At this time, all parties that have responsibilities stated in this data protocol are expected to be aware of this data protocol. It is not expected that the Institutes will directly implement this Policy and Protocol as, as stated in article 2, the Institutes will as a pre-requisite need to develop their own Research Data Management Policy or equivalent document as an elaboration of the present Faculty Policy and Protocol.

The data protocol is reviewed and updated on a regular basis every 2 years, or when changes are made to upstream policies and guidelines that significantly affect the nature or scope of the present FSW protocol.

The data protocol is approved by the Faculty Board of the Faculty of Social and Behavioural Sciences, see article 23.

Appendices

A. Glossary

The glossary presents the following definitions in alphabetical order.

CoreTrust Seal

Core Trust Seal [15] is a certification for repositories engaged in long-term preservation and sharing of research data based on the TRUST principles [16] (Transparency, Responsibility, User focus, Sustainability, Technology). The certification evaluates repositories' technical infrastructure and standards, their organisational, financial, staffing and legal aspects as well as their workflows, risk management,...

De-identification (ano- or pseudonymisation)

Methods to remove all or part of the identifying information or separate the identifying information from the research data in order to make (indirect) identification more difficult.

Data Management Plan (DMP)

Data Management Plans “describe the data that is used and produced during the course of research activities, where the data will be archived, which licenses and constraints apply, and whom credit should be given “ [9].

Data preservation

Data preservation refers in this protocol to storage of research data once research has been performed. It can concern data related to a publication or data in relation to a project that has ended even if they do not directly relate to a publication. In this document, the term data preservation will be preferred to data archiving as it reflects a broader concept including both limited-time preservation for data falling under the WMO or GDPR umbrella or longer-term data archiving.

Data publication

Data publication refers in this protocol to the action of publishing research data either in a repository or in a journal dedicated to publishing data (or software³⁹). Data publication can be implemented with more or less strict conditions depending on the licence chosen and on access conditions (restricted access, embargo periods, etc)

Data storage

Research data storage refers in this protocol to the storage of data while the research project is ongoing.

Metadata

Metadata refer in this protocol to all data about research data that are necessary to understand the research output by humans and machines. Taking the example of a video taken by a researcher, metadata could be the title of the video, the name of its creator, its subject, its language, its licencing (defining who and how it can be reused or not), ... Metadata

³⁹ One example would be the “Journal of Open Source Software” which publishes software in Open Access without any fee for the author or the institution.

are most often saved as “readme files”. Pre-defined metadata standards to create rich metadata exist and can be used, please refer to section 21.2 for more details.

Persistent Identifier (PID)

Persistent identifiers can be defined as long-lasting references to a digital resource. They reliably point to and unambiguously and uniquely identify a digital entity. [17]

Examples of PIDs are

- DOI for articles, datasets, etc;
- ORCID for authors;
- Grant-ID for grants;
- ROR for institutions;
- SWHID for software....

Publication Package

A specification for the preservation of all research material in relation to a publication. FSW has adopted the Publication Packages, based on standards developed in their field by the Deans of Social Sciences of the Netherlands.

Research Data Management (RDM)

RDM refers to the way research data are handled throughout the research life cycle. More specifically, RDM concerns how researchers:

- Create data and plan for its use;
- Organise, structure, and name data;
- Keep data (make it secure, provide access, store and back it up);
- Handle, modify and apply versioning to the data during data analysis;
- Find information resources;
- Share with collaborators;
- Publish data and get cited;
- Sustainably preserve their research data after research.

“Research Data Management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results.” [20]

Research integrity

As defined by the Netherlands Code of Conduct for Research Integrity, research integrity is based on five guiding principles: Honesty, Scrupulousness, Transparency, Independence and Responsibility. For more details, please see [21].

B. Research Data Futures: Preservation and Deletion Policy and Protocol after the Initial Minimal Retention Period

Version: 3.0.

Written by: Céline Richard

Reviewed by (in alphabetical order): Raymond van Erkel, Mitch van Geel, Andrew Hoffman, Katie Hudson, Verena Ly, Kerwin Olfers, Jaap-Willem Mink, Willemijn Plomp

Status: approved by the Faculty Board on 2023-12-11

a. Policy

i. Introduction and purpose

In 2023, three institutes of FSW requested guidance on the destruction of physical and digital data housed on premises of Leiden University or on university network storage. This section outlines the acceptable processes and conditions under which this can be carried out. It should not be interpreted as a requirement for institutes to follow; rather, it simply lays out a clear set of steps for doing so should other institutes wish to carry out similar processes in the future.

The 2014 Dutch Code of Conduct for Scientific Practice followed by Article 14 of the 2021 Data Management Regulations of Leiden University, hereafter referred to as RDM2021, define a minimum retention period of 10 years for all research data. In the absence of a Faculty-level policy framework defining data deletion, research data has accumulated both in physical archives and digital storage systems with detrimental consequences on several levels. Visible consequences include the lack of space for physical storage experienced by several Institutes of the Faculty as well as the rising storage costs for digital data on University network solutions (such as J:/ drive). More broadly, the storage of digital data does not only present financial costs but also environmental costs, adding to the need for limiting the storage of data lacking adequate documentation or of uncertain provenance for undefined periods of time. For some research of the Faculty, not deleting datasets from our storages also constitutes a breach of privacy and trust towards the participants of the studies when agreements explicitly mention that research data will be destroyed as soon as they will not be further used.

By creating a framework for research data management after the mandatory 10 years retention period defined by RDM2021, this document does not aim at alienating researchers from decisions over the future of their research data, but rather at sparking essential conversations over their possible futures. It answers the practical questions voiced by some of FSW's constituent Institutes as well meets existing legal requirements concerning research data as discussed with multiple stakeholders, namely the Documentary Information and Archiving division (DIA, Documentaire Informatievoorziening en Archiefbeheer), the Copyright

Expert of the University Library (UB Auteursrecht Expert) and the Legal Affairs division (Juridische Zaken).

As with other parts of the Faculty Research Data Policy and Protocol, this document aims at being a basis for elaborations by the Institutes. Its impact will be regularly evaluated, and its content will be updated every two years.

i. Scope and audience

This policy concerns research data as defined in the Faculty Research Data Policy and Protocol (see Appendix to this document) that have been preserved for the minimal retention period of 10 years as defined by article 14 of RDM2021. These research data can be:

- non-digital, paper-based, preserved in shared physical archives of the Institutes;⁴⁰
- digital, preserved in University-based infrastructure.

This document applies to all employees and persons performing research under the responsibility of Leiden University (art. 3). This includes external PhD candidates and contract PhD candidates, visiting researchers, retired colleagues and any other guests or partners who carry out research at the Faculty of Social and Behavioural Sciences.

Research conducted by bachelor's and master's students falls under the formal responsibility of their supervisors.

ii. Relevant legislations, agreements

Legislations:

- GDPR (General Data Protection Regulation) applies to any research dealing with Personally Identifiable Data.
- Wet op het Hoger onderwijs en Wetenschappelijk onderzoek (WHW)
- WMO (Wet Medisch-wetenschappelijk Onderzoek met mensen) applies for medical research. This legislation concerns some research projects in the Institutes of Psychology and Education and Child Studies.
- Archiefwet
- Auteurswet

National Guidelines

- 2014 Dutch Code of Conduct for Scientific Practice (Nederlandse Gedragscode Wetenschapsbeoefening) from VSNU (currently Universiteiten van Nederland)
- 2018 Netherlands Code of conduct for Research Integrity from NWO

University guidelines

- 2021 Research Data Management Regulations (Regeling Datamanagement) of Leiden University
- Leiden University policy on privacy and information security

Faculty-specific guidelines and protocols

- Ethics committees guidelines
- Institute-specific protocols

iii. Research data deletion mandate from the Faculty Board

⁴⁰ For non-digital data, we do not consider other data than paper-based data (e.g. biological samples which retention and deletion are ruled by other policies and regulations).

Discussions with relevant University stakeholders, namely the Documentary Information and Archiving division (DIA, Documentaire Informatievoorziening en Archiefbeheer), the Copyright Expert of the University Library (UB Auteursrecht Expert) and the Legal Affairs division, have led to the following conclusions.

There is no legal retention time for research data mentioned within the Archiefwet. The minimal 10 years retention time of article 14 of RDM 2021 rather comes from the 2014 Dutch Code of Conduct for Scientific Practice. For data of medical research projects as defined by the WMO, the minimal retention period is longer and depends on the nature of the research carried out. In the rest of this document, we will thus only mention “minimal retention period” without specifying its duration.

There are no copyrights or intellectual property defined for data by legal texts. Research data constitute an output of the work of researchers hired by Leiden University. Leiden University as the employer therefore has a claim on their ownership. There is no formal requirement to request the researcher’s consent to delete them should it be needed. However, the employer’s consent in the form of a mandate from the Faculty Board to delete research data that have reached the 10 years retention mark is necessary.⁴¹ This mandate has been obtained on the 11th of December 2023 for research data deletion conducted in accordance with this policy and protocol.⁴²

iv. Storage systems review

Physical archives reviews⁴³

The review of the shared physical archives of Institutes falls under the responsibility of the Institute board (or their delegate).

The Faculty requests the Institutes to carry an initial review of the research data within these archives implying the making of a catalogue or register of the preserved research datasets, including at least the year of the start of preservation and the researcher concerned.⁴⁴

When such a review is initiated, the research data management team can help.

Research data that have reached the end of the minimal retention period will be handled according to the research data preservation or deletion process described below.

The institutes will review the content of their archive on an annual basis and make decisions about further preservation or deletion for the datasets that have reached the end of the minimal retention period as described below.

Digital University-based research data storage infrastructure reviews

The responsibility to initiate the review of the University-based research data storage infrastructure lies with the Institute Board (or their delegate). The details of such a review

⁴¹ This mandate concerns only the deletion of research data and not possible transfer of data between Institutions, e.g. if a former Leiden University researcher would like to transfer data to their new Institution – in which case a Data Sharing Agreement is needed.

⁴² This mandate is especially critical for the Education and Child Studies Institute that has on its J:/ drive and in its physical archive research data that are part of the “special category of personal data” described by the GDPR. Some of these datasets have been stored for more than 10 years and according to the GDPR, they should be deleted, however the lack of an existing Faculty mandate prevents the Institute from doing so.

⁴³ This paragraph does not concern CAOS and CWTS that do not have a shared paper archive managed by the Institute.

⁴⁴ The start of the preservation being defined in the main Faculty research data policy and protocol.

such as frequency and systems concerned are elaborated on in the Institute's research data management policy and protocol.

When such a review is initiated, the Faculty Information Manager and ICT coordinator can support with the help of the Institute's research data steward and relevant stakeholders from the Institutes.

Following a review, the research data that have reached the end of the minimal retention period will be evaluated for further preservation or deletion according to the process described below.

v. Research data preservation or deletion process

The research data preservation or deletion process described in this paragraph concerns physical or digital research data that have reached the end of their minimal retention period, as discovered through a storage system review (cf hereabove). The responsibility of this process lies with the research director (or their delegate).

Making reasonable efforts to contact the Principal Investigator

As mentioned in the relevant legislation and agreements paragraph, there is no formal requirement to request the consent of researchers to delete research data that have reached their minimal retention period. However, ethical and research heritage conservation considerations lead the Faculty to request its Institutes to make reasonable efforts to contact researchers concerned before deleting the research data. Should the projects be collaborative, the Principal Investigator is the person to contact preferentially. The Faculty does not expect the Institutes to contact all researchers involved in collaborative projects as it would defeat the notion of reasonable efforts.

Some institutes of the Faculty may choose to contact researchers' relatives should the researcher have passed (unless documents found about the research would prevent to do so, e.g. Informed Consent Forms mentioning that the data should be accessed only by the researcher).⁴⁵

The protocol section of this document presents a 3-step protocol for this purpose.

This contact attempt is not necessary where pre-existing agreements for the storage of the research data exist (e.g. as defined in the framework of an "exitgesprek", DMP or other research documentation).

The notion of reasonable efforts mentioned above means to exclude extreme cases such as, but not limited to, those described below:

- Impossibility to link a dataset to a researcher. Orphaned research data on J:/ drive constitute one example as the creator's name of a dataset cannot be accessed even by the system administrator. If not otherwise documented, attribution of such datasets is thus almost impossible and the Faculty will not ask the Institutes to try to match the concerned research datasets with past researchers.
- Impossibility to find contact details of a researcher no longer working at the University. For example, should remaining staff lists or standard internet search fail to provide contact details of a researcher no longer working for Leiden University and known to

⁴⁵ This does not mean to imply that relatives gain authorship rights over the research but rather aims at offering relatives the possibility in some cases to keep a record of the work of their family member.

have created a concerned research dataset, then the Faculty will not ask the Institutes to make further efforts in contacting them.

Selecting research data to be further preserved or deleted

When the contact attempt is successful, the researcher is asked to go through the datasets and select which research data should be further preserved from those which can be deleted – and carry therefore the responsibility for these decisions.

The research data stewards of the Institutes and the Faculty research data manager can support this process.

If research datasets including personal data are preserved, the data minimisation principle has to be taken into account with the help of the Faculty privacy officer.

Choosing a long-term conservation platform

For the research data that needs to be further preserved, a more appropriate solution than University-based storage systems will be decided on by the researcher within the scope of what is legally possible (if GDPR and / or WMO applies to the dataset).⁴⁶ The research data stewards of the Institutes and the Faculty research data manager can support this process.

In choosing a long-term conservation platform the Faculty encourages researchers to consider:

- The type of platform needed which can take multiple forms depending on the research concerned, ranging from national or discipline-specific repositories (such as DANS services) to formal archives (such as the “National Archief” that could be a relevant choice for some researchers of the Faculty) or even Libraries (University Library, KB,...);
- The sustainability of the chosen platform (certification of a repository, ...);
- The services provided by the platform and their relevance for the considered research dataset (access conditions of the research datasets, documentation (metadata) and cataloging, embargo periods, supported licences, facilitation of data sharing, ...);
- Any relevant legal framework that may limit available options such as the GDPR for research datasets including personal data;
- Any ethical considerations linked to the choice of a conservation platform (e.g. for researchers working with indigenous data, ...).

In the case of orphaned data, i.e. contact attempts were unsuccessful or it was impossible to link a research dataset to a past or present researcher of the concerned Institute, the final preservation or deletion decision lies with the research director or their delegate.

b. Protocol

As mentioned previously, the Faculty requests from its constituent Institutes to elaborate on the content of this policy and protocol as any document developed at the Faculty level cannot properly account for the specificities of the research carried within each Institute. This

⁴⁶ It is possible that in some specific cases the decision of maintaining the storage as initially set is taken. The Faculty asks the Institutes to reflect on these situations in their Institute protocol if they foresee them occurring for their researchers.

protocol section will thus be very minimal and only set a framework for contacting the concerned researchers or their relatives as mentioned in the section “Research data preservation or deletion process”.

3-step protocol for contacting researchers (or their relatives)

- Step 1 consists of sending an initial email to all known email addresses of the person to be reached. Post can be considered should no email address be found. This communication should describe the dataset (as clearly as possible given the status of the documentation found), mention the end of the 10 years minimal retention period, ask for the participation of the person contacted and set a reasonable deadline for them to recontact the sender (this deadline should account for potential unavailability of researchers due to attending conferences, field work in certain research fields of the Faculty, ...).
- Step 2 consists of sending a reminder to the person to be reached 2 weeks before the set deadline.
- Once the set deadline is reached, step 3 consists of sending a third communication to the concerned person informing them that as no answer was received from their side, a final decision about the deletion of the research dataset will be made by the research director of the Institute or their delegate and set a deadline for this decision.⁴⁷

⁴⁷ Setting a final deadline gives to the contacted person a last chance to reconsider their decision.

C. FSW research data storage guiding matrix

		Research drive	Surf drive	Yoda	Leiden Data Store	P-drive	J-departments	J-workgroups	J-research data (J-bulk)	One drive	Teams / Sharepoint	USB device	Dropbox
RDM	Preferred solution		✓	✓	✗	✗	✗	✓	✓	✓	✓		✗
Fit for usage	In development	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
	Pilot	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
	General use	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
	General use with dedicated data manager for the research project	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Data sensitivity	low	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓		!
	Basic	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓		!
	Sensitive	✓		✓	✓	✗	✗			✓	✓		✗
	Critical					✗	✗			✓	✓		✗
Data sharing / users	Single user	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓
	With Leiden researchers	✓	!	✓	✓	✗	✓	✓	✓	!	✓	✗	✓
	With external parties	✓	!	✓	✓	✗	✗	✗	✗	!	!	✗	✓
Data volume	< 50 GB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Between 50 GB and 500 GB	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓
	Between 500 GB and 1 TB	✓	✗	✓	✓	✗	✗	✗	✓	✓		✓	✓
	> 1 TB	!	✗		✓	✗	✗	✗	✓	✗	✗	✓	✗
Complexity	Connection to (lab) devices	✗	✗	✗	✓	✗	✓	✓	✓	✗	✗	✗	✗
	Low latency	✗	✗	✗	✗	!	!	!	!	✗	✗	✓	✗
	High volume read / write	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
	High reliability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
Services	Versioning	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓
	User-friendly	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
	Encryption	✓	✗	✗	✓	✗	✗	✗	✗	✓	✓		?
	Tools for collaboration on data	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✓
Persistence	Remains if the researcher changes affiliation	✓		✓	✓	✗	✓	✓	✓		✓	✓	✓
	Is removed if the researcher changes affiliation	✗		✗	✗	✓	✗	✗	✗	✓	✗	✗	✗
Supported Not supported Contact data steward Consider encryption Not recommended Unsure yet													

D. **Links to Policies or equivalent documents from the Institutes of the Faculty referring to Research Data Management in whole or in part**

Researchers looking for practical information should first look into their Institute's policies / protocols. Guidelines in relation to specific methods used by several Institutes can also be found in the other appendices of the Faculty data protocol.

1. **Centre for Science and Technology Studies**

The Centre for Science and Technology Studies has an [Open Science Policy](#) and [Data Management Guidelines](#).

2. **Cultural Anthropology and Development Sociology**

Cultural Anthropology and Development Sociology has a data management [Policy](#) that includes an annotated Leiden University DMP template and an [appendix about informed consent](#).

3. **Education and Child Studies**

Education and Child Studies and Psychology share common practical resources for data management in relation to writing Data Management Plans and to creating Publication Packages:

- [Annotated Leiden University DMP template on Zenodo](#);
- Set of instructions for [Archiving Publication Packages at the Institutes of Psychology and Education and Child Studies](#).
- Each of these institutes has developed their own Open Science policy that relates partially to data management, but more broadly aims at fostering Open Science. They provide a summary of good research practices that are encouraged and required during the different research stages. Institute of Education and Child Studies – [Open Science Policy and Guidelines](#)
- Institute of Psychology – [Open Science Practices and Guidelines](#)

4. **Political Science**

The Institute of Political Science is currently establishing its first Research Data Management and Open Science Policy and Guidelines. The document has been approved on the 25th of April 2024.

5. **Psychology (please see 3. Education and Child Studies)**

E. Research Data Management infrastructure in development or that would benefit FSW researchers

This section is a request from central level of the University. Infrastructure gaps for Research Data Management exist that seriously affect the daily work of FSW researchers leading them to sometimes choose unsupported solutions. Developing infrastructure is thus necessary but is meaningless without a proper evaluation of the new infrastructure in terms of privacy and information security as well as proper support from product managers at central level.

1. Classification of existing infrastructure

The most urgent need for FSW researchers is an evaluation from the central privacy office and information security of all the storage solutions available for the researchers as well as an interface to communicate on this with researcher. This is still presently missing for almost all infrastructure made available by the University for research data leaving Research Data Management teams in a difficult situation to advice researchers. This evaluation should include the ALICE HPC cluster. Such a classification and a broad communication about it, has been already implemented by other Dutch Universities such as Utrecht (<https://tools.uu.nl/storagefinder/>), Delft (<https://softwarefinder.tudelft.nl/>), Wageningen (<https://library.wur.nl/storagefinder/>).

2. New infrastructure

A list of the most urgently needed infrastructure missing for FSW researchers is:

- Affordable and functional storage during research for very large datasets (>1TB), some of them involving sensitive data as defined by the GDPR;
- Affordable longer-term storage for very large datasets (>1TB) after research some of them involving sensitive data as defined by the GDPR;
- Effective transfer solutions for these very large datasets (including to/from the HPC clusters ALICE and SHARK);
- Secure storage and analysis infrastructure for extra sensitive data irrespective of the dataset size.
- Increased support from central level for the YODA pilot.

Other infrastructure solutions would also benefit FSW researchers but at this stage are not as critical as the points mentioned above.

Bibliography

- [1] “Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>”.
- [2] “<https://www.organisatiegids.universiteitleidennl/en/faculties-and-Institutes/social-and-behavioural-sciences/Institutes/psychology/ethics-committee>”.
- [3] “<https://www.universiteitleidennl/sociale-wetenschappen/pedagogische-wetenschappen/ethiek-commissie>”.
- [4] “<https://www.organisatiegids.universiteitleidennl/en/faculties-and-Institutes/social-and-behavioural-sciences/ethics-review-committee-social-sciences>”.
- [5] “Leonelli S. What Counts as Scientific Data? A Relational Framework. *Philos Sci.* 2015 Dec 1;82(5):810-821. doi: 10.1086/684083. PMID: 26869734; PMCID: PMC4747116.”.
- [6] “<https://researchsoftware.pubpub.org/pub/be6brbxh/release/1>”.
- [7] “Katz DS, Chue Hong NP, Clark T et al. Recognising the value of software: a software citation guide [version 2; peer review: 2 approved]. *F1000Research* 2021, 9:1257 (<https://doi.org/10.12688/f1000research.26932.2>)”.
- [8] “<https://researchsoftware.pubpub.org/>”.
- [9] “Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. *PLoS Comput Biol* 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>”.
- [10] “Annalisa Landi, Mark Thompson, Viviana Giannuzzi, Fedele Bonifazi, Ignasi Labastida, Luiz Olavo Bonino da Silva Santos, Marco Roos; The “A” of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence* 2020; 2 (1-2): 47–55. doi: <https://doi.org/>,” [Online].
- [11] “<https://researchwiki.solo.universiteitleidennl/xwiki/wiki/researchwiki.solo.universiteitleidennl/view/Main/>”.
- [12] “https://www.library.universiteitleidennl/binaries/content/assets/ul2ub/research--publish/cds/rdm-reference-materials/file_naming_and_coding_garp-example_20151211.pdf”.
- [13] “https://www.library.universiteitleidennl/binaries/content/assets/ul2ub/research--publish/cds/rdm-reference-materials/versioning_authenticity_20160318.pdf”.
- [14] “<https://www.openaire.eu/how-do-i-license-my-research-data>”.
- [15] “<https://www.coretrustseal.org/>”.
- [16] “Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>”.
- [17] “Cruz, Maria, & Tatum, Clifford. (2021). NWO Persistent Identifier Strategy (Version 2). Zenodo. <https://doi.org/10.5281/zenodo.4695367>”.
- [18] “<https://gdpr-info.eu/>”.
- [19] “<https://www.staff.universiteitleidennl/ict/privacy-and-data-protection/general-data-protection-regulation-gdpr?cf=social-and-behavioural-sciences&cd=fsw-board-office>”.

- [20] "Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'.
- [21] "Netherlands Code of Conduct for Research Integrity 2018, <https://doi.org/10.17026/dans-2cj-nvwu>".
- [22] "<https://www.rijksoverheid.nl/onderwerpen/rechten-van-patient-en-privacy/medisch-wetenschappelijk-onderzoek>".
- [23] "<https://wetten.overheid.nl/BWBR0009408/2020-01-01>".
- [24] "<https://gdpr-info.eu/>".
- [25] "<https://open.overheid.nl/repository/ronl-dd12795b-eea8-4e23-b552-96ef285cb9ad/1/pdf/Handleiding%20Algemene%20verordening%20gegevensbescherming.pdf>".
- [26] "<https://wetten.overheid.nl/BWBR0005682/2023-01-01>".
- [27] "<https://wetten.overheid.nl/BWBR0007376/2020-01-01>".
- [28] "<https://wetten.overheid.nl/BWBR0001886/2022-10-01>".
- [29] "<https://www.universiteitenvannederland.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf>".
- [30] "<https://www.utwente.nl/en/bms/datalab/datasharing/guideline-faculties-of-behavioural-sciences-def.pdf>".
- [31] "<https://www.organisatiegids.universiteitleidenn.nl/binaries/content/assets/ul2staff/reglementen/onderzoek/research-data-management-regulations-leiden-University.pdf>".
- [32] "<https://www.staff.universiteitleidenn.nl/ict/privacy-and-data-protection>".
- [33] "<https://ichgcp.net/nl>".
- [34] "<https://wetten.overheid.nl/BWBR0005682/2023-01-01>".
- [35] "https://www.nfu.nl/sites/default/files/2021-01/21.00024_Guideline_Quality_assurance_of_research_involving_human_subjects_dec20_0.pdf".
- [36]
- [37] [https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_\(2014\).pdf](https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2014).pdf).
- [38] <https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not>.