This template is a proposal to help FGSE researchers think about a DMP. It contains examples of data management for different types of datasets. If you have other funding sources than SNF and UNIL or that you work with partners outside UNIL, check how this could impact your data management.

For questions, suggestions, feedback, please contact the FGSE research consultant: amelie.dreiss@unil.ch

The DMP can evolve during the project: do not hesitate to contact UNIL service to ask your question about the management of your data, at all stage of your project.

SNF DATA MANAGEMENT PLAN

TABLE OF CONTENTS

1. DATA COLLECTION AND DOCUMENTATION

1.1. WHAT DATA WILL YOU COLLECT, OBSERVE, GENERATE OR REUSE?

A total of [x] datasets, listed below in Table 1, will be generated during the course of the project.

TABLE 1 – Dataset types, format & volume

#	Dataset type	Data format	Volume (in KB, physical space…)	[possibly:] Research objective & Project stage
1	**Raw data:** Samples of …, Recordings … Measures …, Survey, Field notes, Archives, Codes, meeting minutes	Handwritten notebook, specimens, instrument specific format		
2	**Secondary data:** Purchased from…, extracted from … database		Not known at this point	
3	**Processed data:** Analysed data, Simulated data, Interview transcription…, Image of…, Measures of…			
…	**Converted data for storage/analyse:** Scan of Hand-written notes (**Derived from dataset #x**)	.pdf, .jpeg…		
	Final dataset for sharing: …	Pdf/A, mov, csv… (Archiving formats)		
	…			

1.2. HOW WILL THE DATA BE COLLECTED, OBSERVED OR GENERATED?

The research team will follow the guidance from the UNIL Directives 4.2 (Intégrité scientifique) and 4.5 (Traitement et gestion des données de recherche) for research planification and data access.

TABLE 2 – Datasets collection methods, standards & dataset organisation [add possibly in Table: **person/service in charge** for each aspect Collection/Quality/Naming/Organisation]

#	Collection tools / methods	Collection standards	Naming convention[1]	Organisation of datasets
	- Purchased by …, Reused/Download from …, - Survey software, transcripts of semi-structured interviews - Photographed with…, - Measured with… - Electronic Lab Notebook (ELN) - Observation of …	- According to published protocol (ref) - Standard preparation techniques - According to methodology presented in the proposal [part x]	- Identification number for each specimen: xx(project number)_xx(observer)_xx(sample) - File names: Project-Observer-YYYYMMDD-Versionx.format	- Placed in a box labelled Project_place_date - Database (FileMaker…) - File binders…

[Examples of collection details:]

Data collection methods

Materials of dataset #x will be recorded using …

The PACTT has approved [the research contract / the data transfer agreement / the non-disclosure agreement] related to the dataset #x, according to the UNIL Directive 4.1.

Quality assurance processes: standards & controls

Quality and consistency of #x data will be guaranteed through [the calibration of devices, repetition of experiments, repeated measures, appropriate controls, comparison with literature/internal standards/previous data, by verification and confirmation with an expert/peer review…]. [brief description]. Any procedure that differs significantly from those in the published User's Guide will be separately documented. The calibration and reproducibility details of instruments will be stored with the collected data.

Follow-up and peer review of data: regular supervision and lab meetings will be used to ensure that procedures have been carried out correctly and that all data are properly recorded. [brief description]

File organization and versioning

For all generated digital files, we will use a **versioning**[2] method, creating data subsets distinguished by a subscript (v01, v02, etc). All versions of files will be kept.

[1] Check suggestions for naming rules on UNIL Open Science website (in French).
[2] Check suggestions for naming rules on UNIL Open Science website (in French).

Training (techniques & data management) for lab staff will be organized to ensure high quality data. PhD students and postdoctoral researchers will follow continuing education via regular workshops in Open Science, Data security & Data management at the Graduate Campus and at the UNIL Ci.

Organisation: Naming conventions, version control and folder organization will be under the **responsibility** of [x / each work package leader]. Guideline will be communicated to all team members. See section 1.3 for details on dataset organisation and associated metadata.

1.3. WHAT DOCUMENTATION AND METADATA WILL YOU PROVIDE WITH THE DATA?

It will be the **responsibility** of [x] to annotate the data with metadata. The PI will check [weekly during the field season, monthly otherwise…] with all participants to ensure data is being properly processed, and documented.

All the steps described below guarantee the data to be understood by other members of our research group and add contextual value to the dataset should it be reused in the future.

Digital Dataset:

The # datasets will be organized into a database placed in directories, organized as follows.

A **README**.txt, located in the parent folder of Database #x, will describe [the directory hierarchy, its content, …][3].

A « Record-Keeping working paper » describing the workflow (research processes, management of the research project, …) will be placed with the Readme.txt and in SEVAL repository.

Each final dataset will be accompanied by an **INFO**.txt file according to the **metadata standards**[4] (placed in the dataset storage directory and in archive deposit).
[or] There are no agreed metadata standards for our data. We will follow the **informal community best-practice standards** (e.g. ref) of providing all calibration and reproducibility information for our measurements.
The metadata will describe:

- Title:
- Creator: [Last name, first name]
- Date:

[3] Guide to writing "readme" style metadata (Cornell University)
[4] For archived and shared datasets, it is recommended to use **metadata standards** developed in your domain, compatible with your repository and Machine-readable. To find suitable standards, use for instance Research Data Alliance Metadata Standards Directory or DCC Disciplinary Metadata.
Examples of metadata standards:

- Dublin Core Metadata: Interdisciplinary, Dublincore generator
- Datacite Metadata: Interdisciplinary, Datacite metadata generator.
- Access to Geosciences/Biologiacl Collection Databases, ABCD EFG
- ISO 19115: internationally-adopted schema for describing geographic information and services

- Subject: [Keywords]
- Description: [Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data]
- Type: [dataset, image, audio, physical object, etc.]
- Format: [Details of the file format]
- Software/Device, Version, Precision, Calibration
- Identifier: [DOI, …]
- Dataset citation:
- Contact:
- Access rights: [closed access, embargoed access, restricted access, open access.]

Specimens/collection:

The specimens from dataset #x will be placed [in a **box** with a label indicating the specimen number]. The specimen number will be directly written on the specimen (see Table 2).

Each specimen will be entered into the digital **dataset #x** associated with metadata, describing [the locality number, GPS coordinates, date of collection, name of collector, and date of entry into the catalogue, …]. This dataset will be easily searchable.

Images, lab measures:

Each file has associated **embedded internal metadata** produced automatically by the [device] and the computer program used to remote control the apparatus, including details on […]. All data produced would enable another researcher to reproduce images/measures of the specimen under the exact same conditions. Each measure/image will be included into the digital **dataset #x** with associated metadata [on the locality number, GPS coordinates, date of collection, name of collector, sample number, and date of entry into the catalogue].

Codes, models:

A text file will be produced alongside each version of each dataset, indicating [who ran the analysis, date of analysis, and what settings were used during the analysis]. The metadata will be written in text format and associated with each analysis by subject and version number. This allows any researcher to run the exact same analysis under the exact same conditions in the future.

2. ETHICS, LEGAL AND SECURITY ISSUES

2.1. HOW WILL ETHICAL ISSUES BE ADDRESSED AND HANDLED?

No ethical issues:

In the present study, there will be no human subject or human sample involved. There are no ethical issues in the generation of results from this project. All experimental practices will be transparently documented in case any ethical issues do arise.

The data are not subject to any confidentiality agreement, and do not concern any personal details requiring special protection, and no consent needs to be sought.

No permission is required to obtain or process the data.

Authorisations, agreements, contracts :

A non-disclosure agreement / data transfer agreement /… has been drawn up with xx. It has been submitted to the **PACTT office**, which has ensured compliance with legal and ethical principles, such as intellectual property[5].

The Dataset #x has been **obtained** by […], and this **transaction** is documented in […].

All research in this project takes place under a formal agreement between UNIL and [x] and has been **approved by [x ministry, x authority]**. All researchers participating in project fieldwork will obtain **formal permissions** from appropriate [x] authorities each time they travel to the country and go to specific field areas. As such, fieldwork also involves contact and permissions with local authorities.

The drilling operations will **comply with all local and national regulations** applicable at the sites of drilling, including aquifer cleanness and environmental noise.

As the project involves **human subjects**, it will be evaluated by the Commission d'éthique de l'UNIL (CER-UNIL)[6].

As the project completion requires **human cells/tissues** or the use of **sensitive data** related to health, it is in the framework of the LHR (Loi fédérale relative à la recherche sur l'être humain) and an ethical authorization will be obtained from the cantonal ethics committee CER-VD.

[5] According to the Directive 4.1, all UNIL members who undertake a contract must announce their project to the PACTT before the start of negotiations.
[6] This SSP Guide will help you decide whether you need an attestation : https://www.unil.ch/files/live/sites/ssp/files/recherche/Services_chercheurs/CER-GuideFormulaire.pdf

Measures to address ethical issues:

The protocol has been designed according to the ethical charter in the domain, the professional associations or scientific journals recommendations][7].

[E.g. of measures:]

- A storage space for sensitive data will be open at UNIL CI [see wiki-CI]
- Interviewees and focus group participants will be informed about the research and its purpose [possibly explain how].
- Participants will sign a written **Consent form**, as far as possible and according to the circumstances. We will make sure consent forms make provision for future sharing of data.
- Participants must give an oral informed consent in order for the interview/survey to continue.
- All personal data (from interviewees and focus group participants) will be **anonymized**[8] in such a way that it will be impossible to attribute data to specific persons.
- We will not be depositing .wav files as this would compromise that guarantee. However, anonymised transcripts of the interviews and focus groups will be deposited.
- All personal data will be **pseudonymized**. The correspondence table will be encrypted and access restricted to the project leader. All identifying information will be kept in a locked filing cabinet and not stored with electronic files.
- A confidentiality clause may be requested for some interviews with professionals. Depending on the explicit agreement of these interviewees, their anonymised interviews may or may not be made accessible.
- All sensitive **data will be encrypted**[9] and encryption keys will be managed only by [x].
- Field data will be stored on an **encrypted hard drive**.

Measures to address conflict of interest:

The responsible person will inform the Dean of any financial interests that may interfere with their research activities, in accordance with UNIL Directive 4.2 (Scientific integrity).

Conflicts of interest will be declared in the SNSF application, in accordance with Art. 1.18 of the General implementation regulations for the Funding Regulations.

Measures to address individual security issues:

The PI assures that appropriate health and safety procedures conforming to relevant local/national guidelines/legislation are followed for staff involved in this project. Following the Faculty Directive sur la sécurité et les mesures de protection pour le travail de terrain, sous forme de guideline pour une évaluation autonome et personnelle des risques, the health and safety of all participants in the research (investigators, subjects involved or third parties) will be assessed and the necessary security measure will be taken.

[7] e.g., from the International Society of Ethnobiology, Swiss Ethnological Society, …
[8] The Amnesia web application allows you to transform databases into anonymous datasets. It is available free of charge.
[9] Cryptomator, an open source software recommended by UNIL-Ci, is a simple tool to encrypt files

For field work abroad, the research team will follow the guidance the GSE Faculty Directive Travaux de recherche impliquant des déplacements à l'étranger.

2.2. HOW WILL DATA ACCESS AND SECURITY BE MANAGED?

Data access and security management are described in Table 3 below.

[Do not hesitate to contact UNIL Ci at helpdesk@unil.ch before the start of the project to set up tools for sharing and storage security, adapted to your specific project (secured collaborative NAS space, generic account[10]…)]

TABLE 3 – Risk assessment and security measures for data storage and access [add possibly in Table: **person/service in charge** for Storage, Access et Security]

#	Risk assessment	Confidentiality / risk level	Storage methods	Access methods	Security [of data] and protection [of personal information] measures
	- Data loss - Data theft - Data degradation - Security of subjects - Disclosure of sensitive data - …	- Low - Medium - High	- Premises [x] - Laptop of [x] - Server [x] - Too early to specify - …	- Access to all team members without restriction via …, - Access to partners via… - Open access via repository [x] - …	- Securing premises and equipment (computers, storage devices, servers, etc.) via single sign-on, keys, access codes and passwords - Backups of data, at regular intervals and on at least three different media - Regular updating of environments, software and computer programs - Checks on the authenticity and integrity of files - Data encryption - Anonymization or disidentification of confidential or sensitive data

[Examples of access and security management:]

Specimens from Dataset #x: The **storing room** (see section 3.1 for details) has restricted access controlled by key-card entry. Only authorised persons are able to enter the storage area. This secures the specimens from any theft or disturbance.

Datasets #x do not contain confidential or sensitive data, and do not need any additional security measures to protect the data beyond what is delineated in section 3.1.

For dataset #x, we will use UNIL solutions for data **storage** and **backup**, described in section 3.1 below.

Sensitive data management: see section 2.1 above.

[10] More information on UNIL Ci : Comptes génériques

Data **sharing** between team member will be done via SwitchDRIVE and data transfer via SWITCHfilesender (Swich services)[11].

2.3. HOW WILL YOU HANDLE COPYRIGHT AND INTELLECTUAL PROPERTY RIGHTS ISSUES?

The University of Lausanne holds the intellectual property rights to the data generated in this project and in the course of the professional activity as a UNIL collaborator of the project leader, according to the LUL (*Loi du 6 juillet 2004 sur l'Université de Lausanne*)[12] and the Directive 4.5 (*Traitement et gestion des données de recherche, Art. 5*), while the authors of the data retain their copyright.

[Examples of copyright management:]

Our study should not provide concern for commercial and **patenting** issues. If this is not the case, we will contact the UNIL office for technology transfer (PACTT). [If you cannot share particular data because of commercial and patenting concerns, please contact the PACTT and explain the specific constraints.]

This project is being carried out in collaboration with an industrial partner. The intellectual property rights are set out in the **collaboration agreement** [x] approved by the PACTT.

We will promote sharing and unlimited use of the data that we produced using explicit **licences**. For sharing our data, we will use the Creative common licence[13] CC BY - Attribution (so that the data set to get properly cited) or CC BY-NC (if we realise that the data has the potential for commercial purposes and that we intend to perform this development).

[11] If you use another sharing tool, it is recommended to encrypt the sensitive datasets.
[12] Art. 70 LUL. See also Data ownership on UNIL website. This means that for any use other than scientific and academic purposes, UNIL approval is necessary (e.g. for patenting and commercialization). If multiple institutions are involved, please discuss the property issue with your partners.
[13] To choose a licence, see https://creativecommons.org/share-your-work/ or UNIL website on Licence of use.

3. DATA STORAGE AND PRESERVATION

3.1. HOW WILL YOUR DATA BE STORED AND BACKED-UP DURING THE RESEARCH?

Appropriate data storage and back-up is the responsibility of [x].

Specimens/collection:

Specimens from dataset #x: Enough space has been allocated in the [premise] to ensure safe storage under controlled environmental conditions [specify].

Digital Dataset:

Data on computers from dataset #x will be automatically **saved** by UNIL tool[14]. This service is free for UNIL staff.

The dataset #x will be **stored** using Division calcul et soutien à la recherche – DCSR (UNIL Ci) resources[15]. Two copies of the dataset, with volumes of [x GB] will be secured. The estimated direct **cost** will be [...CHF/TB/year, i.e. ...CHF/year][16]. The **Ci service at UNIL provides appropriate protection of data**, including the confidentiality and integrity of the data. The Ci will not modify or transmit the data except in accordance with Swiss law and in cases of assistance and prevention of technical problems. For information security, UNIL follows international standards: ISO 27001 - International information security standard (https://www.itgovernance.co.uk/iso27001).

3.2. WHAT IS YOUR DATA PRESERVATION PLAN?

The data to be preserved long-term after completion are datasets #x-y. We will archive data 10 years after publication according to UNIL standards (Directive 4.2 art. 2.4)[17].

Digital Dataset:

We will deposit dataset #x via our selected open repositories [e.g., SwissUbase, Zenodo, Dryad, Github]. This repository fulfils the requirements from the domain journals and the SNF (non-profit and

[14] See security tools from UNIL Ci: Sauvegarde et sécurité. Special measures to be taken if you use a laptop that you regularly move or if you telework (directive 1.40), such as FileVault (disc encryption for Mac), BitLocker (disc encryption for Windows), VPN crypto (safe internet connection if you are outside the UNIL network).
[15] See UNIRIS page "Storage & security".
[16] see UNIL Ci (DCSR) direct costs 1. These costs can be included in SNF Research Costs: "Computing and storage". See General Implementation Regulation 2.12
[17] "Le chef de projet doit veiller à ce que les données de base produites dans le cadre de la recherche soient conservées en sécurité pendant au moins dix ans après l'achèvement de l'étude. En cas de départ de l'institution, il doit s'assurer d'une conservation appropriée des données."
See also "Preservation and sharing" or LTS on UNIRIS website.

providing FAIR[18] data) and will ensure longevity of the dataset in the long-term. In addition, our chosen data storage facility (see section 3.1) retains primary and secondary research data supporting published articles for at least [# years] after publication.

The estimated **cost** will be [x CHF]. [Zenodo: free, up to 50GB. Dryad: see costs]

[Or:] We will find a data repository appropriate to our domain on a later stage [specify], using re3data.org registry.

For dataset #x, we will use […] **file formats**[19] for data sharing to ensure a file's preservation and re-usability.

Specimens/collections:

As requested by standard practice in the domain, dataset #x, [comprising all material described, figured or mentioned in the text of a scientific publication…] will be deposited in a permanent archive [University, Museum, Library…] that ensures open access of material to future researchers [during x years].

Samples will be stored in the central Institute Facility for at least 10 years after the end of the project.

The estimated **cost** will be [x CHF].

[18] Findable, Accessible, Interoperable, and Re-usable, see explanation on SNF Open research data website.
[19] See recommended formats for data preservation on UNIRIS "preservation and sharing" website.

4. DATA SHARING AND REUSE

[It is possible to include in SNF Research Costs: "Material costs: Costs for granting access to research data (Open Research Data)" see General Implementation Regulation 2.13, generally max=10'000 CHF. For instance for data anonymisation, data organisation, etc.]

4.1. HOW AND WHERE WILL THE DATA BE SHARED?

Repositories making datasets available.

Data from dataset #x will be shared [date / at the time of publication] via the data repository (see section 3.2). Appropriate metadata (see section 1.3) will be associated with the dataset.

Premises making Specimens/collections available

Access to the **Specimens** from dataset #x in the archive (see section 3.2) will be managed by […], [stage/ at the time of publication]. Corresponding metadata (see section 1.3) will be deposited on [repository] in order to allow the collection information to be findable.

Visibility and valorisation of datasets:

Datasets will be given a Digital Object Identifier (DOI) and associated metadata. The DOI corresponding to the datasets in the repository will be included in the article's reference list, allowing identification and access of any dataset in a publication.

The information on datasets (DOI, metadata, repository name) will be uploaded on ***MySNF*** (in 'datasets' section) at the time publication.

We will also link DOIs to appropriate records in the University's publication repository **Serval**[20] [Warning: Serval is not suitable to deposit a dataset], to enhance the dataset's visibility. Metadata about datasets will be publicly searchable and discoverable and will indicate how and on what terms the datasets can be accessed. We will display information about datasets on the lab's webpage, on researcher profile pages **Unisciences**[21], as well as on researchers **ORCID** iDs, which will increase the visibility of the datasets.

We will share data using CC0 or CC BY licences[22] that will become citable products of research (See section 2.3).

[20] Serval user guides and tutorial : https://www.unil.ch/openscience/en/home/menuinst/open-access/serval/how-to-use-serval.html
[21] For a support concerning Unisciences, contact : Floriane Beetschen
[22] https://creativecommons.org/share-your-work/

4.2. ARE THERE ANY NECESSARY LIMITATIONS TO PROTECT SENSITIVE DATA?

[Examples of answers (see also 2.1) :]

This project will **not produce sensitive human personal data** and will not necessitate specific limitations on data sharing. We do not anticipate that this study will generate patentable data or proprietary data that would need protection.

We will restrict data sharing only in cases of concern related to commercial and patenting issues. Except for this, we will share data as widely as possible using Creative Commons licences.

Not all social data will be published. Some interview and observation notes are highly contextual, potentially sensitive, and not appropriate to release. Data deemed highly sensitive and/or confidential will be preserved yet not shared.

Because the survey dataset #x contains **sensitive data**, and the anonymisation is not possible [possibly explain], data will be made available only upon request via [FORS, beQuali, SwissUbase]. The sensitive nature of these data will require that the data be released through a restricted use contract, to which each respondent will give explicit consent.

Because the dataset #x is **confidential** and **of stakeholders' requirements** [explain], the data will be preserved (See section 3.2) but not shared.

4.3. ALL DIGITAL REPOSITORIES I WILL CHOOSE ARE CONFORM TO THE FAIR DATA PRINCIPLES

Yes

4.4. I WILL CHOOSE DIGITAL REPOSITORIES MAINTAINED BY A NON-PROFIT ORGANISATION

Yes

5. ANNEXES

5.1. GUIDELINES

- UNIL Manage your research data
  DMP | Organize your data | Storage & security | Preservation & Sharing

- SNF DMP guidelines, DMP content
- FORS : draft a SNF DMP for social science
- The DMLawTool offers a decision tree to guide researchers through juridic aspects of research data ; it identifies possible solutions to copyright and data protection issues

5.2. RULES

- Article 47 Publication et mise à disposition des résultats de la recherche (Règlement des subsides FNS)
- 2.13 Frais d'exploitation: frais de mise à disposition des données de recherche (Open Research Data) (General implementation regulations for the Funding Regulations)

5.3. TRAINING

- Regular training given by Graduate Campus UNIL: DMP, Open Research Data, Data organisation
- UNIL Ci training on IT security