**Template for the SNSF Data Management Plan**

Introduction

This guide provides instructions for creating a data management plan (DMP) as required by the Swiss National Science Foundation (SNSF) as mandatory part of grant applications.

The SNSF requirement for a DMP serves to implement it’s [policy on Open Research Data](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/). Projects that create a DMP, follow it and keep it up-to-date will be in an excellent position to publish their research data as “Open Data” at the end of the project, which is the main goal of SNSF’s Open Data Policy and can also have a number of advantages for the researchers and institutions involved, e.g. a potentially increased citation rate [[Piwowar2013]](https://opendata.eawag.ch/dmp/References.html#piwowar2013).

The other reason to implement a DMP is a better guard against data loss, better chances for a maintained high data quality, more effcient research organization and a better chance that the reserch project stays within schedule.

A DMP should be regarded as a project management tool and exist as a “living document” that prescribes and tracks all facets of data management during the research data lifecycle. “Data” should be seen as a ressource that needs to be managed, just as the financial budget or human resources.

Guidline and help

* This document was not vetted by the SNSF. It contains answer possibilities from which you can choose if you plan to implement it like that. Do not copy & paste blindy!
* There are section with answer possibilities for openBIS and without openBIS. OpenBIS is an electronic notebook & laboratory information system currently in a pilot phase at Empa. If you are planning to use openBIS you can utilize the sections with openBIS. For more details about openBIS see <https://sis.id.ethz.ch/software/openbis.html#openbis-for-illumina-ngs>
* The structure and questions of the document follow the mySNF form for the data management plan.
* For better overview and jumps between section use navigation: To view the Navigation pane in Word, click the “View” tab in the Ribbon. In the “Show” button group, check the “Navigation Pane” checkbox to turn the Navigation pane on or off. By default, the “Navigation” pane is docked on the left side of the application window.
* For more details and answer examples you can consult the [DLCM template](https://documentation.library.ethz.ch/display/DD/Guidance%2Bfor%2BETH%2Bresearchers%2Bon%2Bfilling%2Bout%2BSNSF%2BData%2BManagement%2BPlans) written by ETH. The template in this document is based on the following documents [DLCM\_SNSF-DMP\_v2.pdf](https://documentation.library.ethz.ch/display/DD/Guidance%2Bfor%2BETH%2Bresearchers%2Bon%2Bfilling%2Bout%2BSNSF%2BData%2BManagement%2BPlans), [SNSF-DMP-openBIS-template.pdf](https://sis.id.ethz.ch/researchdatamanagement/SNSF-DMP-openBIS-template.pdf) (from SIS ETHZ), [Eawag SNSF Data Management Plan Guide](https://opendata.eawag.ch/dmp/index.html) and an Internal DMP from a project
* This template is a work in progress and your feedback is very welcome and also needed for its improvement. Please contact anusch.bachofner@empa.ch (Scientific IT and Research Data Manager).

Contents

[Introduction 1](#_Toc30004919)

[Guidline and help 1](#_Toc30004920)

[1. Data collection and documentation 5](#_Toc30004921)

[1.1 What Data will you collect, observe, generate or reuse? 5](#_Toc30004922)

[1.1.1 What type, format and volume of data will you collect, observe, generate or reuse? 5](#_Toc30004923)

[1.1.2 Which existing data (yours or third party) will you reuse? 5](#_Toc30004924)

[To answer both questions briefly describe the following 5](#_Toc30004925)

[1.2 How will the data be collected, observed or generated? 5](#_Toc30004926)

[1.2.1 What standards, methodologies or quality assurance processes will you use? 5](#_Toc30004927)

[Recommendation for openBIS users 5](#_Toc30004928)

[NOT openBIS users briefly describe the following if planned 5](#_Toc30004929)

[If you are working with personal data describe the following 5](#_Toc30004930)

[1.2.2 How will you organize your files and handle versioning? 6](#_Toc30004931)

[Recommendation for openBIS users 6](#_Toc30004932)

[NOT openBIS users briefly describe the following if planned 6](#_Toc30004933)

[1.3 What documentation and metadata\* will you provide with the data? 6](#_Toc30004934)

[1.3.1 What information is required for users (computer or human) to read and interpret the data in the future? 6](#_Toc30004935)

[1.3.2 How will you generate this documentation? 6](#_Toc30004936)

[1.3.3 What community standards\* (if any) will be used to annotate the (meta)data? 6](#_Toc30004937)

[To answer the questions you can use the following text snippets if you plan to implement it like that: 6](#_Toc30004938)

[2. Ethics, legal and security issues 7](#_Toc30004939)

[2.1 How will ethical issues be addressed and handled? 7](#_Toc30004940)

[2.1.1 What is the relevant protection standard for your data? Are you bound by a confidentiality agreement? 7](#_Toc30004941)

[2.1.2 Do you have the necessary permission to obtain, process, preserve and share the data? Have the people whose data you are using been informed or did they give their consent? 7](#_Toc30004942)

[2.1.3 What methods will you use to ensure the protection of personal or other sensitive data? 7](#_Toc30004943)

[Be aware if you have any personal or otherwise sensitive data: 7](#_Toc30004944)

[Two Possibilities to answer the three Questions above: 7](#_Toc30004945)

[2.2 How will data access and security be managed? 8](#_Toc30004946)

[2.2.1 What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks? 8](#_Toc30004947)

[2.2.2 How will you regulate data access rights / permissions to ensure the security of the data? 8](#_Toc30004948)

[2.2.3 How will personal or other sensitive data be handled to ensure safe data storage and transfer? 8](#_Toc30004949)

[Recommendation for openBIS users (see also below the additions for sensitive data and collaborations outside Empa network) 8](#_Toc30004950)

[Recommendation for NOT openBIS users (see also below the additions for sensitive data and collaborations outside Empa network) 9](#_Toc30004951)

[If you have to share data outside of Empa network describe additionally how access restrictions are implemented and how data security is managed, you may choose some of the following options: 9](#_Toc30004952)

[If you are working with personal or sensitive data you should additionally outline the security measures and describe your security measures, access restrictions, storage and back-up systems 9](#_Toc30004953)

[2.3 How will you handle copyright and Intellectual Property Rights issues? 10](#_Toc30004954)

[2.3.1 Who will be the owner of the data? 10](#_Toc30004955)

[2.3.2 Which licenses will be applied to the data? 10](#_Toc30004956)

[2.3.3 What restrictions apply to the reuse of third-party data? 10](#_Toc30004957)

[To answer the questions consider this and describe accordingly: 10](#_Toc30004958)

[3. Data storage and preservation 10](#_Toc30004959)

[3.1 How will your data be stored and backed-up during the research? 10](#_Toc30004960)

[3.1.1 What is your storage capacity and where will the data be stored? 10](#_Toc30004961)

[3.1.2 What are the back-up procedures? 10](#_Toc30004962)

[Recommendation for openBIS users (\*) 10](#_Toc30004963)

[3.2 What is your data preservation plan? 11](#_Toc30004964)

[3.2.1 What procedures would be used to select data to be preserved? 11](#_Toc30004965)

[To answer the question consider this and describe accordingly: 11](#_Toc30004966)

[3.2.2 What file formats will be used for preservation? 11](#_Toc30004967)

[To answer the question consider this and describe accordingly: 11](#_Toc30004968)

[4. Data sharing and reuse 11](#_Toc30004969)

[4.1 How and where will the data be shared? 12](#_Toc30004970)

[4.1.1 On which repository do you plan to share your data? 12](#_Toc30004971)

[4.1.2 How will potential users find out about your data? 12](#_Toc30004972)

[Answer options: 12](#_Toc30004973)

[Important fact: SNSF does not pay for storage in commercial data repositories (Please don't forget in your SNSF petition to check the offered 10'000 CHF) 12](#_Toc30004974)

[4.2 Are there any necessary limitations to protect sensitive data? 12](#_Toc30004975)

[4.2.1 Under which conditions will the data be made available (timing of data release, reason for delay if applicable)? 12](#_Toc30004976)

[Answer options: 12](#_Toc30004977)

[4.3 All digital repositories I will choose are conform to the FAIR Data Principles 13](#_Toc30004978)

[4.4 I will choose digital repositories maintained by a non-profit organisation 13](#_Toc30004979)

1. Data collection and documentation

## 1.1 What Data will you collect, observe, generate or reuse?

### 1.1.1 What type, format and volume of data will you collect, observe, generate or reuse?

### 1.1.2 Which existing data (yours or third party) will you reuse?

#### To answer both questions briefly describe the following

* **categories of datasets** you plan to generate or use
* **their role in the project**
* **reference of the source** (if relevant) for reused data
* **estimation of volume** of raw and curated data
* **format (preferably open standard)** of raw and curated data (which device is used to create/simulate/download the data)
* **e.g.**
	+ 1. Category: reaction parameters … (with their role, reference, volume and format)
	+ 2. Category: images …
	+ ….
	+ Or at the end summary of role, reference, volume and format of the data

## 1.2 How will the data be collected, observed or generated?

### 1.2.1 What standards, methodologies or quality assurance processes will you use?

#### Recommendation for openBIS users

All experimental data will be automatically imported into the institutional electronic Laboratory Information System (LIMS) from the measurement device. Methods and materials will be recorded using the institutional Electronic Lab Notebook (ELN).

#### NOT openBIS users briefly describe the following if planned

For each dataset in your project (including data you might re-use) mention:

* the use of core facility services (specify their certifications, if any),
* the use of standards or internal procedures; describe them briefly.
* the use of appropriate controls to ensure data quality; describe them briefly. (e.g. peer review, comparison with literature/internal standards/previous data, plausibility checks, repetition of experiments, range check, double data entry, statistical or visual outlier detection, instrument verification tests, etc.)

#### If you are working with personal data describe the following

* Have the subjects of your data collection (persons) been fully informed (what data do you collect, what will you do with the data, and who will receive it, when will it be deleted) and have the subjects given their informed consent?
* Have the subjects of your data collection (persons) been informed about their rights on information, data deletion and data correction?

### 1.2.2 How will you organize your files and handle versioning?

#### Recommendation for openBIS users

All files produced during this project will be stored in our Electronic Laboratory Notebook (ELN) and Laboratory Information Management System (LIMS) openBIS. In this ELN, each scientist has a personal folder where projects and experiments are organized. Each experiment is described in the electronic notebook and all data related to the experiment is directly attached to it, in so called “datasets”. Each dataset is immutable, thus different file versions are stored in the lab notebook in different datasets with a manually generated version number. Very large datasets (100s of TBs) are not directly stored in openBIS datasets, but they

are linked to the experimental description using an extension to openBIS called BigDataLink. This works similarly to the git version control software, so every time changes are made to the data, these need to be committed to openBIS, which automatically keeps track of the versioning.

#### NOT openBIS users briefly describe the following if planned

* used naming convention, i.e. the structure of folders and file names you will use to organize your data. For example: Project-Experiment-Scientist-YYYYMMDD-HHmm-Version.format (concretely: Atlantis-LakeMeasurements-Smith-20180113-0130-v3.csv)
* used form of database
* used version control system, such as Git.

## 1.3 What documentation and metadata\* will you provide with the data?

\*Metadata refers to “data about data”, i.e., it is the information that describes the data that is being published with sufficient context or instructions to be intelligible for other users. Metadata must allow a proper organization, search and access to the generated information and can be used to identify and locate the data via a web browser or web based catalogue.

### 1.3.1 What information is required for users (computer or human) to read and interpret the data in the future?

### 1.3.2 How will you generate this documentation?

### 1.3.3 What community standards\* (if any) will be used to annotate the (meta)data?

\*look for metadata standards here: [Research Data Alliance Metadata Standards Directory](http://rd-alliance.github.io/metadata-directory/standards/)

#### To answer the questions you can use the following text snippets if you plan to implement it like that:

* ***Depending on your procedure choose one of the following:***
	+ ***For openBIS users***: In the data management system (openBIS ELN-LIMS), metadata are provided as attributes of the respective datasets. Based on the defined metadata schema, openBIS ELN-LIMS will be configured so that the required metadata is automatically assigned to datasets and / or manually provided by the researcher.
	+ ***You are not an openBIS user and plan to use existing metadata standards***: The data will be accompanied by the following contextual documentation, according to [name of the standard practice] for [name of the project]
	+ ***You are not an openBIS user and there are no existing metadata standards:*** The metadata that would best describe the data depends on the nature of the data. For research data generated in project [name of the project], it is difficult to establish a global criteria for all data, since the nature of the initially considered data sets will be different, so that the metadata will be based on a generalised metadata schema as the one used in Zenodo (Zenodo is a public repository in which the data of a publication can be made open accessible, which is a requirement for SNSF, view section 4.1) .
* ***Next snippet not depending on openBIS:***

Each data set file (spreadsheet or CSV) will be accompanied by a text document (README file) explaining in detail the methods/software used to generate the data, and the necessary information to interpret the data. The metadata will have the following information included:

• Title: free text,

• Creator: Last name, first name,

• Date,

• Subject: Choice of keywords and classifications,

• Description: Text explaining the content of the data set and other contextual information, needed for the correct interpretation of the data,

• Format: Details of the file format,

• Resource Type: data set, image, audio, etc.,

• Identifier: DOI,

• Access rights: closed access, embargoed access, restricted access, open access.

• Information about the methodology used, the performed processing and analytical steps, variable defintions, references to vocabularies used, as well as units of measurement.

2. Ethics, legal and security issues

## 2.1 How will ethical issues be addressed and handled?

### 2.1.1 What is the relevant protection standard for your data? Are you bound by a confidentiality agreement?

### 2.1.2 Do you have the necessary permission to obtain, process, preserve and share the data? Have the people whose data you are using been informed or did they give their consent?

### 2.1.3 What methods will you use to ensure the protection of personal or other sensitive data?

#### Be aware if you have any personal or otherwise sensitive data:

* Make sure you have a good idea what personal data means in the context of Swiss data protection law: Personal data refers to any information that relates to a particular person. Next to items such as physical or email-address, health record, or age, it also includes for example photographs, videos recordings or genetic information. If in doubt, consult Empa Legal and/or the Empa Ethics Commission.
* Check whether your project involves the collection, storage or processing of personal data.
* Check whether your project involves other types of sensitive information, e.g. otherwise not easily obtainable information about vulnerabilities of water infrastructure or locations of rare and protected species.
* Check whether your work involves data that you obtain under a non-disclosure agreement or any kind of contract that would restrict its usage or dissemination, or includes other special requirements relevant for data handling.
* If your project requires the assessment of the internal Ethics Commission (or similar), or requires permission from third parties such as a cantonal office, check whether any imposed requirements are related to data collection, processing or dissemination.

#### Two Possibilities to answer the three Questions above:

1. If you consider that there are **no ethical or confidentiality issues** in your project, you can use the following statement:

There are no ethical or confidentiality issues in the generation of results from this project.

1. If you consider that **there are** **ethical or confidentiality issues** in your project, explain how these issues will be managed:
* Specify all data that are affected by any of the above points. Specify the special requirements regarding data collection, handling and dissemination.
* Specify that the necessary ethical authorizations will be obtained and from the internal Ethics Commission and/or the Cantonal Ethics Commission.
* The collection of personal data most likely requires informed consent. Describe the content of the consent form and ancillary measures to ensure informedness, if applicable.
* Mention relevant approvals and permissions obtained from third parties and reference their requirements.
* Personal/sensitive data should be anonymized. Describe the anonymization method (e.g. pseudonymization or removal of personal information).
* Access to personal/sensitive data should be restricted. If you use encryption and/or if you put in place special access restrictions, just mention that here and refer to Section 2.2, where you give the details.
* Personal/data should be stored in a secure and protected place. Mention non-technical procedures that ensure data protection, such as scheduled deletion of data or training activities. For technical implementation details or purely technical measures reference Section 2.2.
* Protective measures should be taken with regard to the transfer of data and sharing of data between partners.
* Specify that sensitive data is not stored in cloud services (e.g. data related to individuals, data under a non-disclosure agreement, data injuring third party rights or legal expertise).

## 2.2 How will data access and security be managed?

### 2.2.1 What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks?

### 2.2.2 How will you regulate data access rights / permissions to ensure the security of the data?

### 2.2.3 How will personal or other sensitive data be handled to ensure safe data storage and transfer?

#### Recommendation for openBIS users (see also below the additions for sensitive data and collaborations outside Empa network)

*The project is not expected to generate personal or other sensitive data* ***(write this only if you have NO sensitive data, otherwise replace with section below)****.* Data will be processed and reviewed internally within the project before being made public. All data generated in the project will be stored in our open-BIS ELN-LIMS. This operates in a client-server model, which is installed and maintained by Empa IT or ETH Zurich IT services on Empa or ETH Zurich infrastructure. Researchers can access openBIS via any of the most common web browsers. openBIS requires user authentication with Empa credentials and it provides user right management, so that different users can have different access to all or different parts of the system, as required. Below is a description of the default openBIS roles, which can be modified upon request:

1. Instance admin. Has full admin powers. Can customize settings, create, modify and delete entities, assign user roles, create data spaces.

2. Instance observers. Has read-only access to everything in openBIS.

3. Space admin. Can create, modify, delete entities and assign roles only within a given data space.

4. Space power user. Can create, modify and delete entities only within a given data space.

5. Space user. Can create and modify entities only within a given data space.

6. Space observer. Has read-only access limited to a given data space.

openBIS does not offer any specific option for sensitive data, but the data will be encrypted prior to upload to openBIS. Furthermore, all operations on the system (incl. which users log in and when) are logged, so that it is fully transparent who did what to the data and when. The data stored in openBIS is physically located on a NAS (network attached storage) at Empa or provided by the ETH Zurich IT Services. Data security and confidentiality are protected by using Empa's Microsoft Active Directory authentication. The shared filesystem can only be accessed from inside the Empa or ETH network and remote access is possible by establishing a Virtual Private Network (VPN) that is secured by 2-factor authentication.

#### Recommendation for NOT openBIS users (see also below the additions for sensitive data and collaborations outside Empa network)

*The project is not expected to generate personal or other sensitive data* ***(write this only if you have NO sensitive data, otherwise replace with section below)****.* Data will be processed and reviewed internally within the project before being made public. All data will be stored on Empa’s central shared Fileservices. Data security and confidentiality are protected by using Microsoft Active Directory authentication. The shared filesystem can only be accessed from inside the Empa network and remote access is possible by establishing a Virtual Private Network (VPN) that is secured by 2-factor authentication.

#### If you have to share data outside of Empa network describe additionally how access restrictions are implemented and how data security is managed, you may choose some of the following options:

* Data access rules will be detailed in before starting the project.
* Access to the data/database will be logged, thus each access is traceable.
* Access to laboratory and offices will be restricted to authorized persons. The list of authorized persons will be managed by …
* All servers will be located in a datacentre with restricted access. The datacentre is based in [country] (preferably data are stored at Empa).
* No data will be stored on a public cloud / cloud hosted outside Switzerland.
* As the project partners are located in different countries, the only practical means of sharing data is through the internet. This will be done through the "polybox" cloud service offered to members of the ETH domain, which allows the users to restrict access to those involved in the project, i.e. the applicants and the persons filling the proposed temporary research positions.

#### If you are working with personal or sensitive data you should additionally outline the security measures and describe your security measures, access restrictions, storage and back-up systems

Briefly mention the datasets that require special protection (reference Section 2.1) and use an adverb to indicate the “level of risk” (e.g. “high”, “medium”, “low”).

**Security measures: anonymization / encryption you may choose some of the following options:**

* All personal data will be anonymized in such a way that it will be impossible to attribute data to specific persons.
* All personal data will be pseudonymized. The correspondence table will be encrypted and access restricted to the project leader.
* All sensitive data will be encrypted and encryption keys will be managed only by authorized employees.
* Sensitive data transfers will be end-to-end encrypted.

**Regarding access rights you may choose some of the following options:**

* Sensitive data will be accessible only by authorized participants to the project. The list of authorized participants will be managed by …

**Regarding storage and back-up you may choose some of the following options:**

* All damaged media containing sensitive data will be physically destroyed.
* No sensitive / personal data will be stored in cloud service external to Empa. Sensitive data can be for example data related to individuals, data under a nondisclosure agreement, data injuring third-parties rights or legal expertise.
* All computers storing or computing sensitive data will not be connected to the Internet.
* All computers storing or computing sensitive data will have a hardened configuration (disk encryption, restricted access to privileged accounts to a small, controlled group of users, restricted or disabled remote access using privileged accounts, disabled guest or default accounts, local firewall, automatic screen lock with password protection, disabled remote out-of-band management (IPMI, Active Management Technology (AMT), etc.), disabled USB ports, removable privacy filter on screens, automatic updates via “Windows Update”, Apple’s “Software Update” or Linux “yum auto-update”, anti-virus software, Adobe’s “Flashplayer” and “Java” runtime).

## 2.3 How will you handle copyright and Intellectual Property Rights issues?

### 2.3.1 Who will be the owner of the data?

### 2.3.2 Which licenses will be applied to the data?

### 2.3.3 What restrictions apply to the reuse of third-party data?

#### To answer the questions consider this and describe accordingly:

* If the **data is suitable for sharing**, you could write:

It is observational data (hence unique) and could be used for other analyses. Reuse opportunities are vast. For this reason, we aim to allow the widest reuse of our data and will release them under Creative Commons CC0.

* If there are **more restrictions for sharing the data**, consider the following:

When copyright is applicable, Creative Commons licenses are recommended. However, Creative Commons licenses are not recommended for software. For further questions or in case your data or source code might be commercially exploitable (e.g. patentable), please consult the [Technology Transfer Office at Empa](https://www.empa.ch/web/s607/technologietransfer).

* + If you are working with an industrial partner, this is an example:

This project is being carried out in collaboration with an industrial partner. The intellectual property rights are set out in the collaboration agreement. The intellectual property generated from this project will be fully exploited with help of the [Technology Transfer Office at Empa](https://www.empa.ch/web/s607/technologietransfer). The aim is to patent the final procedure and then publish the work in a research journal and to publish the supporting data under an open Creative Commons Attribution (CC BY) license.

3. Data storage and preservation

## 3.1 How will your data be stored and backed-up during the research?

### 3.1.1 What is your storage capacity and where will the data be stored?

### 3.1.2 What are the back-up procedures?

#### Recommendation for openBIS users (\*)

Our data will be stored in openBIS ELN-LIMS. OpenBIS uses a postgres database that stores all metadata. This database is backed up (“pg\_dump”) every night with a 7 days retention of the dumps and fully backed-up twice a week with a backup retention of 20 days. The full backup procedure includes a point-in-time recovery that allows a finer granularity (up to minutes) of data recovery in case of a disaster. The database backup is stored on the NAS (network attached storage) at Empa or provided by the ETH Zurich IT services. The same NAS is used to store the data uploaded to openBIS. This network attached storage is snapshot every night with a 7 days retention, and data is backed up on a proprietary tape library with a retention of 90 days. Data which is no longer actively needed is moved to the long term storage (i.e. tapes). The tape library where openBIS moves the data has a read-only replica in a different geographical location in order to minimize any data loss.

\*For data linked to openBIS with the BigDataLink tool, please provide

details of the data location and back-up.

***NOT openBIS users can use this snippet if data is stored in Empa local network***

Data will be stored on back-upped servers in the Empa local network. For file services and virtual server farm, Empa shares a server/storage platform (Netapp Metrocluster, Cisco UCS Server, VMWare) with Eawag. The backup procedure is fully automatic. Snapshots of files are taken at least three times during a working day. All data are mirrored synchronously between the two server sites on the Empa-Eawag campus in Dübendorf. Additionally, backups (to disk) are taken from the Metrocluster at a third location on the campus. Backups are kept for three months. We have arranged to have access to the required storage-capacity.

## 3.2 What is your data preservation plan?

### 3.2.1 What procedures would be used to select data to be preserved?

#### To answer the question consider this and describe accordingly:

Note that preservation does not necessarily mean publication (e.g. personal sensitive data may be preserved but never published), but publication means generally preservation.

* Answer options for preservation procedure:
	+ You preserve all relevant data generated or used by the research. This includes all raw data, all processed data that directly underlies the reported results, and all ancillary information necessary to understand, evaluate, interpret and re-use the results of the study. Data from intermediate steps of the analysis that can be recreated from preserved information does not need not to be stored.
	+ Specify which data will be preserved after the completion of the project and the corresponding data selection procedure (e.g. long-term value, potential value for re-use, obligations to destroy some data, anonymization (if necessary), etc.)
* It is necessary to outline a long-term preservation plan for the datasets beyond the lifetime of the project. Mention which datasets are long-term preserved and how it is done. For example after specifying which data you can write:
	+ Data will be preserved long term by depositing to an online repository, and a copy will be kept on servers of the original owner host institution.

### 3.2.2 What file formats will be used for preservation?

#### To answer the question consider this and describe accordingly:

* Comment on the choice of file formats. For example you can write:
	+ Where possible, we will store files in open archival formats e.g. Word files converted to PDF-A or simple text files encoded in UTF-8 and Excel files converted to CSV. In case this is not possible, we will include information on the software used and its version number.

4. Data sharing and reuse

## 4.1 How and where will the data be shared?

### 4.1.1 On which repository do you plan to share your data?

### 4.1.2 How will potential users find out about your data?

#### Answer options:

#### Important fact: SNSF does not pay for storage in commercial data repositories (Please don't forget in your SNSF petition to check the offered 10'000 CHF)

* Check whether there is a well-recognized, specialized data repository for the kind of data you are producing. Mention it if you do so. To check for a repository in a given discipline it is best to search in the Registry of Research Data Repositories ([*https://www.re3data.org/*](https://www.re3data.org/) ). It can be browsed by discipline, data type and country for discovering an appropriate home for your data or to find shared datasets to use in your research. Don't forget to check if the repository fulfills the requirements of SNSF (FAIR principles etc.)
* There is currently no data repository available that is specific to the data. If such a repository becomes available and conforms to the FAIR Data Principles, we will use it. In the meantime, data generated within the project will be stored in the generalist repository XY, which is very flexible in terms of file types and metadata and allows unrestricted access. We choose XY because it supports the FAIR principles and implements long-term preservation features, notably bitstream preservation.

(if you are looking for a generalized repository, which fulfills the requirements of SNSF, use this overview: [*http://www.snf.ch/SiteCollectionDocuments/FAIR\_data\_repositories\_examples.pdf*](http://www.snf.ch/SiteCollectionDocuments/FAIR_data_repositories_examples.pdf))

## 4.2 Are there any necessary limitations to protect sensitive data?

### 4.2.1 Under which conditions will the data be made available (timing of data release, reason for delay if applicable)?

#### Answer options:

* If there are **no ethical, confidentiality or IP issues** and you are **confident that you can make all relevant data public in time**, you can write:
	+ We expect no limitations with respect to publishing the data. It will be made available to the public in full, latest at the time of publication of the project report.
* If there **are ethical, confidentiality or IP issues:**
	+ State how you will deal with them (anonymization etc.)
* If parts of the data will not be made available at all, state the reason(s).
* If you expect delayed publication state the reason(s) for that delay. State explicitly *when* exactly you plan to publish it though.
	+ Possible reasons for delayed publication (**accepted by SNSF**) could for example include:
		- The time necessary to anonymize personal data.
		- The need to keep patentable information secret until patent protection applies.
	+ Possible reasons for delayed publication, which are currently **not accepted by SNSF**, include for example:
		- The intent to synchronize the publication of the data with other publications (e.g. project report, paper, press release) to maximize visibility and impact.
		- The intent to base follow-up publications on the data, after the project has finished.
		- The intent to couple re-use of the data by other groups to an offer for collaboration.

## 4.3 All digital repositories I will choose are conform to the FAIR Data Principles

If your chosen repository is conform to the FAIR Data Principles (see [examples](http://www.snf.ch/SiteCollectionDocuments/FAIR_data_repositories_examples.pdf)) then you can just check it 

## 4.4 I will choose digital repositories maintained by a non-profit organisation

The SNSF supports the use of non-commercial repositories for data sharing. Costs related to data upload are only covered for non-commercial repositories.

-> If the answer is no: “Explain why you cannot share your data on a non-commercial digital repository.”

Otherwise you answer yes.