DLCM template for the
SNSF Data Management Plan

prepared by

ETH *zürich*

ETH Library

EPFL

Please note
Recommendations in this document are intended to illustrate the guidelines and other information provided by the SNSF for preparing Data Management Plans[1]. The SNSF's guidelines are binding.

This document was prepared jointly by teams from the libraries of EPFL and ETH Zurich, with input from Data Life Cycle Management (DLCM)[2] partners, and exists in adapted versions for the two universities. It can also be freely adapted to other institutions' needs. The examples therefore do not cover all disciplines. Further examples from other subject areas and other feedback are welcome to **info@dlcm.ch** for possible inclusion in future revisions.

 For additional resources and tools, please feel free to explore the Swiss DLCM website: **https://www.dlcm.ch**

[1] **http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx**

[2] **http://www.dlcm.ch**

DLCM Version 1.0

License Creative Commons CC BY-SA

mandated by

Table of Contents

DLCM template for the SNSF Data Management Plan

Institution

Responsibilities

Principal Investigator:
(Specify name and email)

Data management plan contact person:
(Specify name and email)

1. Data collection and documentation

1.1 What data will you collect, observe, generate or re-use?

Questions you might want to consider
- What type, format and volume of data will you collect, observe, generate or reuse?
- Which existing data (yours or third-party) will you reuse?

Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated datasets.

This relates to the FAIR Data Principles F2, I3, R1 & R1.2
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Recommendations

For each dataset in your project (including data you might re-use) mention:
- **Data type:** Briefly describe categories of datasets you plan to generate or use, and their role in the project.
- **Data origin:** to be mentioned if you are reusing existing data (yours or third-party one). Add the reference of the source if relevant.
- **Format of raw data** (as created by the device used, by simulation or downloaded): open standard formats should be preferred, as they maximize reproducibility and reuse by others and in the future [see *List of recommended file formats**].
- **Format of curated data** (if applicable): open standard formats should be preferred [see *List of recommended file formats**].
- **Estimation of volume of raw and curated data.**

* See for example the list of recommended file formats for **EPFL** and for **ETH Zurich**.

Examples of answers
to be adapted to your research application

EXAMPLE 1
The data produced from this research project will fall into two categories:
1. The various reaction parameters required for optimization of the chemical transformation.
2. The spectroscopic and general characterization data of all compounds produced during the work.

Data in category 1 will be documented in [file format].
  Spectroscopic data in category 2 will be produced as [file format] and converted to [file format] for further use. Other characterisation data in this category will be collected in [file format].
  We anticipate that the data produced in category 1 will amount to approximately 10 MB and the data produced in category 2 will be in the range of 4 – 5 GB.

EXAMPLE 2

This project will work with and generate three main types of raw data.

1. Images from transmitted-light microscopy of giemsa-stained squashed larval brains.
2. Images from confocal microscopy of immunostained whole-mounted larval brains.
3. Western blot data.

All data will be stored in digital form, either in the format in which it was originally generated (i.e. Metamorph files, for confocal images; Spectrum Mill files, for mass spectra with results of mass spectra analyses stored in CSV files; TIFF files for gel images; MariaDB SQL dump files for genetics records), or will be converted into a digital form via scanning to create tiff or jpeg files (e.g. western blots or other types of results).

Measurements and quantification of the images will be recorded in excel files (for long term preservation, they will be converted in CSV files. Micrograph data is expected to total between 100GB and 1TB over the course of the project. Scanned images of western blots are expected to total around 1GB over the course of the project. Other derived data (measurements and quantifications) are not expected to exceed 10MB.

EXAMPLE 3

The data are health records auto-generated by users of the application X. They are subjected to a contract with the company X.

All fields contain user observations and entered manually, except for temperature which is measured by a Bluetooth connected thermometer.

Data fields per user (anonymized by X):
• User identifier
• Age
• Weight, Size.

Data fields per users per day of observation:
• Temperature and time at which temperature is taken,
• Cervical fluid quality (none, sticky, creamy, egg white, watery) and quantity (little, medium, lots)
• Cervix height (low, med, high), cervix openness (closed, med, open), cervix firmness (firm, med, soft),
• Sexual intercourse (protected or unprotected),

• Menstruation (light, medium, heavy), spotting, starting a new cycle,
• Custom data (notable predetermined fields are pregnancy test results or ovulation test results).

Data will be received in CSV format, and consists of the record of 2 million users. It will amount to maximum 1GB.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

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

Questions you might want to consider
- What standards, methodologies or quality assurance processes will you use?
- How will you organize your files and handle versioning?

Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc.

Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures.

This relates to the FAIR Data Principle R1
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Recommendations

What standards, methodologies or quality assurance processes will you use?
For each dataset in your project (including data you might re-use) mention:
- the use of core facility services (specify their certifications, if any),
- whether you follow double blind procedures (define it),
- the use of standards or internal procedures; describe them briefly.

If you are working with persons' data, confirm 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 they 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?

How will you organize your files and handle versioning?
Indicate and describe the tools you will use in the project.

You may rely on the following tools depending on your needs:
- a **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)
- a **code revision management system**, such as **Git**. Several Git servers are available for ETH domain: **c4science.ch**, **gitlab.epfl.ch**, **gitlab.ethz.ch**.
- a **data management system**, such as an Electronic Laboratory Notebook / Laboratory Information System (ELN / LIMS). Within ETH domain, examples of used ELN / LIMS: **openBIS**, **SLims**.

Examples of answers

to be adapted to your research application

EXAMPLE 1

The reaction conditions will be recorded and collated using a spreadsheet application and named according to each generation of reaction as follows: ProjectW-ReactionX-GenerationY-ScientistZ-YYYYMMDD-HHmm.csv

The various experimental procedures and associated compound characterization will be written up using the Royal Society of Chemistry standard formatting in a Word document, each Word document will also be exported to PDF-A. The associated NMR spectra will be collated in chronological order in a PDF-A document.

EXAMPLE 2

All samples on which data are collected will be prepared according to published standard protocols in the field [cite reference]. Files will be named according to a pre-agreed convention. The dataset will be accompanied by a README file which will describe the directory hierarchy.

Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information.

Microscope images capture and store a range of metadata (field size, magnification, lens phase, zoom, gain, pinhole diameter etc.) with each image.

This should allow 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.

EXAMPLE 3

Experiments will include appropriate controls to ensure validity [brief description]. Data consistency will be assessed by comparing repeated measures.

EXAMPLE 4

Quality of analytical data will be guaranteed through calibration of devices, repetition of experiments, comparison with literature/internal standards/previous data, by a peer review.

EXAMPLE 5

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).

EXAMPLE 6

The experimental records and observations are recorded by hand-written notes followed by digitization (scanning). The analytical data are collected by the instruments that generated them; they are processed by the native programs [please specify program name, version and file format] associated with the instruments. A periodic quality control process will be applied to remove errors and redundancies. Errors include for example incorrect handling and machine malfunction. The quality control process will be documented.

The quality of experimental records and observations will be controlled by repeating experiments.

For NMR and X-ray, the data collection is done through instrument standardised data acquisition programs. For E-chem, UV-Vis, IR, GC, GC-MS, lab-standardized protocols will be used.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

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

Questions you might want to consider
- What information is required for users (computer or human) to read and interpret the data in the future?
- How will you generate this documentation?
- What community standards (if any) will be used to annotate the (meta)data?

Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data. Metadata should at least include basic details allowing other users (computer or human) to find the data. This includes at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data.

Furthermore, the documentation may include details on the methodology used, information about the performed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement.

Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information.

This relates to the FAIR Data Principles I1, I2, I3, R1, R1.2 & R1.3
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Recommendations

Indicate all the information required in order to be able to read and interpret the data (context of data) in the future. General documentation of the data is often compiled into a plain text or **markdown** README file. These formats may be opened by any text editor and are future proofed.

In addition, for each data type
- Provide the metadata standard used to describe the data (for concrete examples see: **Research Data Alliance Metadata Standards Directory**).

  If no appropriate (discipline oriented) existing standard is available, you may describe the *ad hoc* metadata format you will use in this section.

  Metadata* may also be embedded in the data (e.g. embedded comments for code). Or, when for example using Hierarchical Data Format **HDF5**, arbitrary machine readable metadata can be included directly at any level.
- Describe:
  - the **software** (including its **version**) used to produce the data and the software used to read it (they can be different),
  - the format and corresponding filename extension and its version (if possible).

  The used software should be archived along with the data (if possible, depending on the software license).
- Describe the automatically generated metadata, if any.
- Provide the data analysis or result together with the raw data, if possible.

Additional information that are helpful in a README file
- description of the used **software**,
- description of the used **system environment**,
- description of relevant **parameters** such as:
  - geographic locations involved (if applicable)
  - all relevant information regarding production of 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.

Examples of answers

to be adapted to your research application

EXAMPLE 1

The data will be accompanied by the following contextual documentation, according to standard practice for synthetic methodology projects:

1. Spreadsheet documents which detail the reaction conditions.
2. Text files which detail the experimental procedures and compound characterization.

Files and folders will be named according to a pre-agreed convention YXZ, which includes for each dataset, identifications to the researcher, the date, the study and the type of data (see section 1.2).

The final dataset as deposited in the chosen data repository will also be accompanied by a README file listing the contents of the other files and outlining the file-naming convention used.

EXAMPLE 2

Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively. It will be responsibility of:

• each researcher to annotate data with metadata,
• the Principal Investigator to check weekly (during the field season, monthly otherwise) with all participants to assure data is being properly processed, documented, and stored.

EXAMPLE 3

IFS and OpenIFS model integrations will be run and standard meteorological and computing performance data output will be generated. Both will be run at ECMWF, and only performance data will be made available to the public. The meteorological output will be archived in MARS, as it is standard research experiment output. The data will be used for establishing research and test code developments, and will enter project reports and generally accessible publications. The IFS will not be made available, OpenIFS is available through a dedicated license.

IFS meteorological output (incl. metadata) and format follows the World Meteorological Organization (WMO) standards. Compute performance (benchmark) output will be stored and documented separately. Data will be in ASCII and maintained locally. The output will be reviewed internally, and the ECMWF facilities allow reproduction of this output if necessary.

EXAMPLE 4

Two types of metadata will be considered within the frame of the project [name of the project]: that corresponding to the project publications, and to the published research data.

In the context of data management, metadata will form a subset of data documentation that will explain the purpose, origin, description, time reference, creator, access conditions and terms of use of a data collection.

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, which includes elements such as:

• 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.

Additionally, a readme.txt file could be used as an established way of accounting for all the files and folders comprising the project and explaining how all the files that make up the data set relate to each other, what format they are in or whether particular files are intended to replace other files, etc.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

DLCM
Research Data
Management Hub

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

Questions you might want to consider
- What is the relevant protection standard for your data? Are you bound by a confidentiality agreement?
- 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?
- What methods will you use to ensure the protection of personal or other sensitive data?

Ethical issues in research projects demand for an adaptation of research data management practices, e.g. how data is stored, who can access/reuse the data and how long the data is stored. Methods to manage ethical concerns may include: anonymization of data; gain approval by ethics committees; formal consent agreements. You should outline that all ethical issues in your project have been identified, including the corresponding measures in data management.

This relates to the FAIR Data Principle A1; see
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_ translation_SNSF_logo.pdf

Recommendations

Description and management of ethical issues
Describe which **ethical issues** are involved in the research project (for example, human participants, collection/use of biological material, privacy issues (confidential/sensitive data), animal experiments, dual use technology, etc.).

For more information, feel free to contact your institution. See for example: The EPFL Research Ethics website:
https://research-office.epfl.ch/ethical-legal-review

The ETH Zurich Guidelines for Research Integrity:
https://doi.org/10.3929/ethz-b-000179298

The ETH Zurich Compliance Guide:
https://rechtssammlung.sp.ethz.ch/Dokumente/133_en.pdf

Explain how these ethical issues will be managed, for example:
- The necessary ethical authorizations will be obtained from the competent ethics committee.
- Informed consent procedures will be put in place.
- Personal/sensitive data will be anonymized.
- Access to personal/sensitive data will be restricted.
- Personal/data will be stored in a secure and protected place.
- Protective measures will be taken with regard to the transfer of data and sharing of data between partners.
- 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.

Ethical authorizations
If your project involves **human subjects**, an ethical authorization from either the cantonal ethics commission or the institutional ethics commission is needed. This depends on whether your project is invasive/non-invasive and whether or not health-related data is collected/used.

- For research involving work with **human cells/tissues**, a description of the types of cells/tissues used in the project needs to be provided, together with copies of the accreditation for using, processing or collecting the human cells or tissues.

- Research which involves the **collection or use of personal data** needs to be reviewed by the cantonal ethics commission.
- If **animal experiments** are conducted in the context of the research project, an authorization of the cantonal veterinarian office is needed.
- **Dual-purpose technologies** (civil and military): Transfer of knowledge, software, demonstrators or prototypes could fall under the scope of the Federal Law on the control of dual-purpose goods (LCB) and its Ordinance (OCB) in the context of technology transfer or research proposals, but also in informal personal contacts. Before transmission of information, research results, prototypes etc. to a company, person or institution (even academic) outside of Switzerland, it must be checked whether the data/information to be transmitted are apt to authorization.
- Research that may have a **negative impact on the environment**, for example research with Genetically Modified Organisms (GMO), requires an authorization from the Federal Office for the Environment (**FOEN**). If the research project has a negative impact on the **health and safety of the researchers** involved (for example if the research proposal involves the use of elements that may cause harm to humans), authorizations for the processing or possession of harmful materials must be requested. For more information, feel free to contact your institution.

Examples of answers

to be adapted to your research application

EXAMPLE 1

Please check if your project involves one of the following ethical issues:

- Human participants (This includes all kinds of human participation, incl. non-medical research, e.g. surveys, observations, tracking the location of people)
- Human cells/tissues
- Human embryonic stem cells
- A clinical trial
- The collection of personal/private data
- Animal experimentation
- Third countries (access and benefit sharing)/export law issues.
- Environmental and/or health and safety issues (for example, a negative impact on the environment and/or on the health and safety of the researchers.)
- The potential for military applications (dual-use technology).

If you consider that there are no ethical issues in your project, you can use the following statement:

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

EXAMPLE 2

This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and correlates of service utilization for these disorders in a nationally representative sample of over 8000 respondents. 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. An ethical authorization will be obtained from the cantonal ethics committee for this project.

EXAMPLE 3

Research in this proposal involves the use of animals of the species mouse (Mus Musculus). Animal studies will be preceded by multiple biochemical experiments in vitro and in cultured cells. Mouse experiments will only be used at advanced stages of investigations when few, specific and highly relevant questions can be addressed by a limited number of experiments.

The PI and the research team will work in conformity with all applicable rules, guidelines and principles such as the EU directive 2010/63/EU on the protection of animals

used for scientific purposes, the Swiss federal law on animal protection (RS 455), the federal ordinance on animal protection (RS 455.1), and the federal ordinance on animal experimentation, production, and housing (RS 455.163). All animal experiments will only be initiated after having received the approval of the Cantonal and Federal authorities.

Details on animal usage

In performing the experiments, we strive to strictly adhere to the 3Rs principle of Replacement, Refinement, and Reduction.

- **Reduction:** Each experiment will be designed to use the minimum number of mice required to obtain statistical significance. For the proposed pharmacokinetic experiments, a total number of 24 mice will be required.
- **Refinement:** The animals will be housed in the animal facilities of ETH, which meet international housing norms, and the animal health status is monitored by a certified veterinarian. To reduce stress and discomfort of the animals, all procedures will be performed only after animals are anaesthetized. After experiments animals will be euthanized. Also, as soon as animals show signs of severe discomfort and / or tumor burden during experiments, they will be euthanized by cervical dyslocation after being anaesthetized.
- **Replacement:** Alternatives for mouse experiments will be considered at all stages during the project. Whenever possible, these alternatives will replace the mouse experiments.

Training

All researchers and technicians working with the animals receive proper animal welfare training in conformity with DFE Ordinance 455.109.1 on 'Training in animal husbandry and in the handling of animals'.

EXAMPLE 4 ....

Environmental protection and safety

The PI assures that appropriate health and safety procedures conforming to relevant local / national guidelines / legislation are followed for staff involved in this project. The health and safety of all participants in the research (investigators, subjects involved or third parties) must be a priority in all research projects (see **http://securite.epfl.ch/page-34437-en.html**). The project will be conducted under EPFL's Lex 1.5.1 – Directive concerning occupational health and safety (DSST). The present directive determines the assignment of functions relating to health and safety in the workplace. It specifies the responsibilities of all the

actors who must work as part of a network at EPFL. It also forms an integral part of risk management, at both CEPF and EPFL levels.

EXAMPLE 5 ....

All data are anonymized, and as such, we are in line with the Swiss Federal Act on Data Protection as described on the page of the **Federal Data Protection and Information Commissioner** (Préposé Fédéral à la Protection des Données et à la Transparence, PFPDT).

EXAMPLE 6 ....

The project respects all the constraints and requirements as laid down in the Swiss Federal Act on Data Protection and supervised by the **Federal Data Protection and Information Commissioner**.

Indeed, as the finality of the project does not relate to individuals and the published results do not allow to identify the participants nominatively, we have communicated with all participants giving them the following basic information:

- Author / Responsible person
- lType and extent of collected / processed data
- Objectives of the processing
- Any communication to be made to third-parties / recipients categories / planned transborders communications, with all necessary guaranties related to **the article 6 of the Data Protection Law**.
- The facultative nature of the participation to this project and the possibility to resign at all timesThe consequences, if any, in case of refusal to participate (No inconvenience should result)
- Access and correction rights

EXAMPLE 7 ....

The project is a medical research project and respects all the rules and regulations laid down in the Swiss Federal Act on Data Protection and supervised by the **Federal Data Protection and Information Commissioner**. We are only using and processing data for individuals who have given their explicit consent.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

2.2 How will data access and security be managed?

Questions you might want to consider
- What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks?
- How will you regulate data access rights / permissions to ensure the security of the data?
- How will personal or other sensitive data be handled to ensure safe data storage and transfer?

If you work with personal or other sensitive data you should outline the security measures in order to protect the data. Please list formal standards which will be adopted in your study. An example is ISO 27001-Information security management. Furthermore, describe the main processes or facilities for storage and processing of personal or other sensitive data.

This relates to the FAIR Data Principle A1
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Recommendations

The main concerns regarding data security are data availability, integrity and confidentiality, in particular the levels of risks involved and technical and organizational measures as named in the Swiss Federal Act on Data Protection.

Define whether:
- the level of the data availability risk is: low / medium / high.
- the level of data integrity risk is: low / medium / high.
- the level of data confidentiality is: low / medium / high.

You may choose some of the following options:

Regarding anonymization / encryption:
- 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:
- Sensitive data will be accessible only by authorized participants to the project. The list of authorized participants will be managed by …
- 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 …

Regarding storage and back-up:
- All data will be backed-up on a regular basis and access to backup media will be
- managed according to data access rules. Backups will be stored in another location.
- All damaged media containing sensitive data will be physically destroyed.
- All servers will be located in a datacentre with restricted access. The datacentre is based in [country] (preferably data are stored at your institution).

DLCM
Research Data Management Hub

- No data will be stored on a public cloud / cloud hosted outside Switzerland.
- No sensitive / personal data will be stored in cloud service external to EPFL / ETH Zurich. "Sensitive data can be for example data related to individuals, data under a non-disclosure 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).

Please note
In May 2018, the EU General Data Protection Regulation (GDPR, Regulation (EU) 2016 / 679) will come into force. This already now influences future cooperation with any EU-based partners and will be implemented in Swiss law, as well.

GDPR introduces an approach of "Privacy by Design" for parties working with personal or other sensitive data, requiring projects to define their data protection measures from the beginning.

Where the GDPR applies you must outline in a Data Protection Impact Analysis (DPIA, text or table, **see example**) the risks involved to the rights of your studies' subjects and the security measures foreseen in order to protect the data. This is crucial for your project. The less risks you have, the better. The more data safeguards you can imply, the better. The earlier stage you imply them at, the better.

Cf. Art 35 of the EU General Data Protection Regulation entering into force May 2018
http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/ ?uri=CELEX:32016R0679&from=EN#d1e3265-1-1

Examples of answers
to be adapted to your research application

EXAMPLE 1
The data will be processed and managed in a secure non-networked environment using virtual desktop technology.

EXAMPLE 2
All interviewees and focus group participants will sign a Consent form agreed to by the School ethics committee. We have guaranteed anonymity to our interviewees and focus group participants. Therefore 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. We will make sure consent forms make provision for future sharing of data. All identifying information will be kept in a locked filing cabinet and not stored with electronic files.

EXAMPLE 3
Data will be stored on the centralized file storage system managed by our institutional or school [please specify, e.g. life sciences or basic sciences] IT department. The access to the data is managed through the ETH identity management system, which is a secured system following the best practices in terms of identity management. Our central storage facility has redundancy, mirroring and is monitored.

Contact for assistance
Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

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

Questions you might want to consider
- Who will be the owner of the data?
- Which licenses will be applied to the data?
- What restrictions apply to the reuse of third-party data?

Outline the owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated including the licence(s). For consortia, an IPR ownership agreement might be necessary. You should comply with relevant funder, institutional, departmental or group policies on copyright or IPR. Furthermore, clarify what permissions are required should third-party data be re-used.

This relates to the FAIR Data Principles I3 & R1.1
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_
translation_SNSF_logo.pdf

Recommendations

Attaching a clear license to a publicly accessible data set allows other to know what can legally be done with its content. When copyright is applicable, **Creative Commons** licenses are recommended. However, Creative Commons licenses are not recommended for software.

Amongst all Creative Commons licenses, CC0 "no copyright reserved" is recommended for scientific data, as it allows other researchers to build new knowledge on top of a data set without restriction. It specifically allows aggregation of several data sets for secondary analysis. Several data repositories impose the CC0 license to facilitate reuse of their content.

In order to enable a data set to get cited, and therefore get recognition for its release, it is recommended to attach a CC-BY "Attribution" license to the record, usually a description of the dataset (metadata). To get recognition, data sets can be cited directly. However, to increase their visibility and reusability, it is recommended to describe them in a separated document licensed under CC BY "Attribution", such as a data paper or on the institutional repository.

When the data has the potential to be used as such for commercial purposes, and that you intend to do so, the license CC BY-NC allows you to keep the exclusive commercial use.

Reuse of third-party data may be restricted. If authorised, the data must be shared according to the third party's original requirement or license.

Examples of answers
to be adapted to your research application

EXAMPLE 1
The research is not expected to lead to patents. Other Intellectual Property Rights (IPR) issues will be dealt in line with the institutional recommendation. As the data is not subjected to a contract and will not be patented, it will be released as open data under Creative Commons CC0 license.

EXAMPLE 2
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 from

DLCM
Research Data Management Hub

the institutional Technology Transfer Office (TTO). 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.

EXAMPLE 3

Data is suitable for sharing. They are observational data (hence unique) and could be used for other analyses or for comparison for climate change effects among many things. 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.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

3. Data storage and preservation

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

Questions you might want to consider
- What is your storage capacity and where will the data be stored?
- What are the back-up procedures?

Please mention what the needs are in terms of data storage and where the data will be stored.

Please consider that data storage on laptops or hard drives, for example, is risky. Storage through IT teams is safer. If external services are asked for, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data.

Please specify your back-up procedure (frequency of up-dates, responsibilities, automatic/manual process, security measures, etc.)

Recommendations

Ask for your institutional storage solutions.

Examples of answers
to be adapted to your research application

EXAMPLE 1
Storage and back up will be in three places:
- On Laptop of [Name of Researcher]
- On institutional collaborative storage
- Other (please specify)

[Name of Researcher] will be responsible for the storage and back up of data. This will be done weekly. Backups on the institutional infrastructure are automated using the RSYNC tool.

EXAMPLE 2
Original notebooks and hardcopies of all NMR and mass spectra are stored in the PI's laboratory. Additional electronic data will be stored on the PI's computer, which is backed up daily. Additionally, the laboratory will make use of the PI's lab server space at institution's storage facility for a second repository of data storage. The PI's lab has access to up to 1 terabyte of information storage, which can be expanded if needed.

All the project data will be stored using the institution's Collaborative Storage, which is backed-up on a regular basis.

EXAMPLE 3
All our data will be uploaded to our Electronic Laboratory Notebook (please specify which one, and where it is ins-talled). The data is stored on institutional storage facilities and it is set up by our IT support to be automatically backed up daily.

EXAMPLE 4
The EPFL centralized file storage service follows the best practices and standards regarding storage, for instance high availability, multiple levels of data protection, partnership with providers for support. The service is managed cen-trally by the hosting department of the Vice Presidency for

Information Systems (VPSI) and ensures security, cohe-rence, pertinence, integrity and high-availability.

Two distinct storage locations can be found on the EPFL campus with replication between the two. Physical servers' pairing and clustering guarantees local redundancy of data. Moreover, volume mirroring protects data in case of disaster on the primary site. The copy is asynchronous and automatic and runs every two hours.

The file servers are virtualized for separation between logical data and physical storage, RAID groups ensure phy-sical storage protection: data is split in chunks written on many disks with double parity. Moreover, volume snapshots are used and can allow user restoration of previous versions if need be. For specific needs, optional backup on tape can also be done.

Access to the data is managed by the owner of the vo-lumes through the identity management system of EPFL. Any person who needs access to data has therefore to be a registered and verified user in the identity management system.

Our data is going to be stored on this platform.

Contact for assistance

Feel free to contact directly your institution to find out more infor-mation regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

3.2 What is your data preservation plan?

Questions you might want to consider
- What procedures would be used to select data to be preserved?
- What file formats will be used for preservation?

Please specify which data will be retained, shared and archived 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, etc.). Please outline a long-term preservation plan for the datasets beyond the lifetime of the project.

In particular, comment on the choice of file formats and the use of community standards.

Recommendations

Describe the procedure, (appraisal methods, selection criteria …) **used to select data to be preserved**. Note that preservation does not necessarily mean publication (e.g. personal sensitive data may be preserved but never published), but publication means generally preservation.

This section should answer the following questions:
- What data will be preserved in the long term – **selection criteria**, in particular:
  - **Reusability of the data:** quality of metadata, integrity and accessibility of data, license allowing reuse, readability of data (chosen file formats)
  - **Value of the data:** indispensable data, completeness of the data or data set, uniqueness, possibility to reproduce the data in the same conditions and at what cost, interest of the data, potential of reuse
  - **Ethical considerations**
  - **Stakeholders requirements**
  - **Costs:** additional costs that come for depositing data in a repository or data archive of your choice (costs anticipation and budgeting)
- Selection basically has to be done together with or by the data producer or someone else with deep specialist knowledge.
- What data curation process(es) will be applied, i.e.: anonymization (if necessary), metadata improvement, format migration, integrity check, measures to ensure accessibility.
- Data retention period (0, 5, 10, 20 years or unlimited)
- Decision to make the data public
- Use of sensitive data (i.e. privacy issues, ethics, or intellectual property laws)
- Definition of the responsible person for data (during the process of selection and after the end of the project)

Other **criteria from the Digital Curation Center** (UK). In addition, select appropriated preservation formats (see section 1.1) and data description or metadata (see section 1.3).

Examples of answers

to be adapted to your research application

EXAMPLE 1

Data will be stored for a minimum of three years beyond award period, per funder's guidelines. If inventions or new technologies are made in connection data, access to data will be restricted until invention disclosures and/or provisional patent filings are made with the institutional Technology Transfer Office.

EXAMPLE 2

We will preserve the data for 10 years on university's servers and also deposit it in an appropriate data archive at the end of the project (e.g. Zenodo, see section 4.1 below). 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.

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

4. Data sharing and reuse

4.1 How and where will the data be shared?

Questions you might want to consider
• On which repository do you plan to share your data?
• How will potential users find out about your data?

Consider how and on which repository the data will be made available. The methods applied to data sharing will depend on several factors such as the type, size, complexity and sensitivity of data.
    Please also consider how the reuse of your data will be valued and acknowledged by other researchers.

This relates to the FAIR Data Principles F1, F3, F4, A1, A1.1, A1.2 & A2
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_
translation_SNSF_logo.pdf

Recommendations

It is recommended to **publish data in well established** (or even certified) domain specific **repositories**, if available:
• **re3data** is a repository directory allowing to select repositories by subject and level of trust (e.g. certifications)
• ETH Zurich researchers are encouraged to publish data in ETH's own **Research Collection** repository to ensure full compliance with ETH regulations.

In domains for which no suitable subject repositories are available, generalist repositories are available.
Among the most common used:
• **Zenodo** (free, maximum 50 GB/dataset, hosted by CERN)
• **Dryad** (120$ for the first 20 GB and 50$ for additional GB, Non-profit organization)
• **Figshare** (free upload, maximum 5 GB/dataset, commercial company)

Note
SNSF does not pay for storage in commercial data repositories (even though data preparation costs are eligible). Check the **SNSF's criteria for non-commercial repositories** (section 5.2). If you choose a commercial repository, read carefully the Terms of service to check if they respond to your needs and to your institutions' ones as well as to your institutional (data) policy.

In order to **make your data findable by other users**, it is important that:
• each data packet and publication has a DOI (or similar persistent identifier) assigned,
• they are deposited Open Access in a repository harvested by the main data services (e.g.: **OpenAire**, **EUDAT**, …).

Examples of answers

to be adapted to your research application

EXAMPLE 1

Some of the ongoing data will be shared on [Researcher]'s Github repository (results and code from the project, data from twitter searches). Major revisions of this page will be baked up using the Github-Zenodo connection (see **https://guides.github.com/activities/citable-code/**). All other data we will be published on Zenodo under CC0 license.

We chose Zenodo because it supports the FAIR principles (**http://about.zenodo.org/principles/**). The immediate publication at the end of the project aims to minimize the data loss risk, while the 2 years embargo guarantees us to be first to exploit our data. Zenodo implements long-term preservation features, notably bitstream preservation.

EXAMPLE 2

Datasets from this work which underpin a publication will be deposited in the ETH Zurich Research Collection, and made public at the time of publication. Data in the repository will be stored in accordance with funder's data policies. Files deposited in the Research Collection will be given a Digital Object Identifier (DOI). The retention schedule for data will be set to 10 years from date of deposition in the first instance, with possible extension for datasets which remain in regular use.

The DOI issued to datasets in the repository can be included as part of a data citation in publications, allowing the datasets underpinning a publication to be identified and accessed.

Metadata about datasets held in the Research Collection will be publicly searchable and discoverable and will indicate how and on what terms the dataset can be accessed.

EXAMPLE 3

For this project, the National Geoscience Data Centre (NGDC) (see **http://www.bgs.ac.uk/services/ngdc/home.html**) is the most suited repository. As it is adapted to geodata, it facilitates storage and allows interactive geographical search. In addition, many other researchers in our field are familiar with it.

This repository requires the deposition under Open Governement Licence (see **http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/**), which demands attribution when the data is reused (our dataset must by cited, similarly to the CC BY license).

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

4.2 Are there any necessary limitations to protect sensitive data?

Questions you might want to consider
- Under which conditions will the data be made available (timing of data release, reason for delay if applicable)?

Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output.

Restrictions may be only due to legal, ethical, copyright, confidentiality or other clauses.

Consider whether a non-disclosure agreement would give sufficient protection for confidential data.

This relates to the FAIR Data Principles A1 & R1.1
http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Recommendations

You may mention specifically the conditions under which the data will be made available:
- there are no sensitive data
- the data are not available at the time of publication
- the data are not available before publication
- the data are available after the embargo of …
- the data are not available because of the patent of … for a period of …

Examples of answers
to be adapted to your research application

EXAMPLE 1
Data which underpins any publication will be made available at the time of publication.

All unpublished data will be deposited in a data repository 12 months after the end of the award.

EXAMPLE 2
Astronomical data will be diffused but under an embargo of one year for priority of exploitation reasons.

EXAMPLE 3
Personal data will be anonymized before diffusion based on the recommendations from the Federal Act on Data Protection (FADP) (see **https://www.admin.ch/opc/en/classified-compilation/19920153/index.html**). The package SDC-Micro (**https://cran.r-project.org/package=sdcMicro**) will be used to assess the risk of identification: we will make sure that each data set has a k-anonymity of 3 at least.

Contact for assistance
Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

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

[CHECK BOX]

Recommendations

The SNSF requires that repositories used for data sharing are conformed to the FAIR Data Principles. For more information, please refer to the **SNSF's explanation of the FAIR Data Principles**.

You can find certified repositories in **Re3data.org**, an exhaustive registry of data repositories.

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

[RADIO BUTTON yes / no]

Recommendations

If you do not choose a repository maintained by a non-profit organization, you have to provide reasons for that.

One possible reason would be to ensure the visibility of your research, if your research community is standardly publishing data on a well-established but commercial digital repository.

Please note that the SNSF supports the use of non-commercial repositories for data sharing. Costs related to data upload are only covered for non-commercial repositories. Check the **SNSF's criteria for non-commercial repositories** (section 5.2).

Contact for assistance

Feel free to contact directly your institution to find out more information regarding this section. In case there are no resources available, feel free to contact the DLCM Swiss initiative via email at **info@dlcm.ch** or by filing out the contact form on the DLCM website: **https://www.dlcm.ch**

5. External useful resources

List of useful tools prepared by the Swiss DLCM project
https://www.dlcm.ch

Digital Curation Centre glossary
http://www.dcc.ac.uk/digital-curation/glossary

Casrai dictionary
http://dictionary.casrai.org/Category:Research_Data_Domain