www.snf.ch
Wildhainweg 3, P.O. Box 8232, CH-3001 Berne

Data Management Plan (DMP) – Guidelines for researchers

1. Introduction

Managing and sharing research data as openly as possible is one of the principles of good scientific practice. The SNSF adheres to this principle, as stated in Article 47 of its Funding Regulations: in stating that *"[…] grantees are obliged to make available to the public in an appropriate manner the research results obtained with the help of SNSF funding, […]"*. The SNSF has set out the criteria it expects funded researchers to meet in its Open Research Data Policy statement. For the implementation of these principles, the SNSF favours a bottom-up approach. It provides best practice guidelines and gives each scientific community sufficient flexibility in defining and applying its own standards. In particular, the best way of managing and sharing data depends on the research field.

The aim of a Data Management Plan (DMP) is to plan the life cycle of data. It offers a long-term perspective by outlining how data will be generated, collected, documented, shared and preserved. A project's DMP can refer to discipline-specific practices and standards. The SNSF provides a template to help researchers complete their data management plan.

The SNSF requires that researchers share at least the data underlying their publications. It expects this data to be shared as soon as possible, but at the latest together with the relevant scientific publication. Data can be raw or processed, depending on the project and the discipline. Datasets must always be carefully documented with associated metadata, such that other researchers understand how the data was collected, as well as under which conditions and how it can be re-used. If specific tools are needed to re-use the data, this needs to be documented and, if possible, the tools made available. In any case, the provided data and documentation (metadata) must be sufficient to ensure their reusability. Furthermore, the information and data provided should allow researchers to reproduce published results, replicate statistical analyses and validate conclusions presented in the scientific publications.

Data sharing – best practices

To facilitate the discovery, access, re-use and citation of datasets, it is important that the publication of research data follows a set of clearly defined and broadly applicable best practices. The FAIR Data Principles[1] define a range of qualities a published dataset should have in order to be Findable, Accessible, Interoperable and Reusable (see SNSF explanation of the FAIR Data Principles). The SNSF expects researchers to share their data according to the FAIR Data Principles on publicly accessible, digital and non-commercial repositories. It is important to note that the FAIR Data Principles do not require researchers to share all their data without any restrictions. Rather they advocate applying a standard procedure when sharing research data for reuse, so that humans and computer systems can easily find, interpret and use them under clearly defined conditions. The FAIR Data Principles are being adopted by a growing number of research funding organisations (e.g. Horizon 2020, NIH).

2. The SNSF's Data Management Plan

To implement its Research Data Policy the SNSF requires information on the life cycle of data at the time when a grant application is submitted. In order to account for different data management practices between

[1] Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016)

disciplines, the SNSF has defined minimum standards for the structure and content of the information to be provided.

2.1. How to submit a DMP

While completing their grant application on mySNF, researchers will be asked to provide information regarding their data management. The DMP form comprises four sections: (1) data collection and documentation, (2) ethics, legal and security issues, (3) data storage and preservation, and (4) data sharing and reuse. Sub-questions and online help texts will help researchers to complete the form (see DMP content).

The DMP is an integral part of the grant proposal. The proposal can only be submitted once the DMP has been completed.

Some research projects do not produce or reuse any data. If this is the case, applicants do not have to complete the whole DMP form. However, they are asked to explain why they do not expect to generate or reuse any data in their proposed research.

Some data cannot be shared because applicants are bound by legal, ethical, copyright, confidentiality or other clauses. They will be asked to explain their specific constraints.

2.2. Assessment of the DMP

The submitted DMP is considered a notice of intention. Its content is assessed by the SNSF Administrative Offices for its plausibility and adherence to the SNSF policy on open research data. It is not part of the scientific evaluation process. Members of the Research Council or Evaluation Panel have access to the DMPs, but will not evaluate these documents. DMPs are not sent out for external review.

Submission of a plausible DMP is a requirement for any transfer of funding. If there are shortcomings in the submitted information provided by the applicants, they will receive a “task” in mySNF to complete/amend specific sections of the DMP at the time of the funding decision.

2.3. Lifetime management

The DMP remains editable during the entire lifetime of the grant. Its contents can be adapted as the project evolves.

In any case, researchers will be prompted to update their DMP at the end of the grant. This updated version will be assessed together with the final scientific report. The SNSF Administrative Offices retain the right to request additional information and/or amendments to the contents of the final DMP.

The final version of the DMP will be made available on the SNSF’s P3 database. This will increase the visibility and impact of the research outcomes by making it easier for other researchers to access and reuse the datasets.

3. Examples of data management plans

DMPs are very individual. They can be of various types and their composition can differ. The examples provided by the Digital Curation Centre (UK) show this diversity.

4. Eligible Costs

The costs of enabling access to research data that was collected, observed or generated under an SNSF grant are eligible if the research data is deposited in recognised scientific, digital data archives (data repositories) that meet the FAIR principles and do not serve any commercial purpose (IR 2.13). It is permissible to upload data to commercial repositories but only the data preparation costs will be covered by the SNSF.

5. Examples of repositories that comply with the FAIR Data Principles and are non-commercial

Requesting that researchers apply the FAIR Data Principles in every detail is an ambitious policy. In addition, finding the “perfect” repository providing all necessary features to host FAIR data can be a challenge. To make the transition towards FAIR research data easier, the SNSF decided to define a set of minimum criteria that repositories have to fulfil to conform with the FAIR Data Principles.

Four repositories which accept datasets from different research fields and fulfill the SNSF requirements are shown (see examples of data repositories). It is, of course, possible to archive data on other (field-specific) repositories. Researchers can proceed as follows to ensure that the chosen repository is in line with the SNSF requirements (non-commercial, FAIR Data Principles).

5.1. Non-commercial repositories

The first step is to consult www.re3data.org, where most repositories are listed.

- Under the tab "Institutions", check if a commercial entity is involved in 'general' or 'technical' responsibility (categories "Type of institution" and "Type(s) of responsibility")
- If not, SNSF considers the repository to be non-commercial (even if 'funding' or 'sponsoring' is provided by a commercial entity).
- If yes, the SNSF considers the solution to be a commercial repository (see Identification of non-commercial repositories).

If the repository is not listed on www.re3data.org, the repository should be contacted to clarify this point. Researchers should also suggest that the repository be included in www.re3data.org

5.2. Repository complying with the FAIR Data Principles

Researchers should check if the repository is compatible with the FAIR Data Principles. The answer to each of the questions below must be "yes" (see examples of data repositories).

- Are datasets (or ideally single files in a dataset) given globally unique and persistent identifiers (e.g. DOI)?
- Does the repository allow the upload of intrinsic (e.g. author's name, content of dataset, associated publication, etc.) and submitter-defined (e.g. definition of variable names, etc.) metadata?
- Is it clear under which licence (e.g. CC0, CC BY, etc.) the data will be available, or can the user upload/choose a licence?
- Are the citation information and metadata always (even in the case of datasets with restricted access) publicly accessible?
- Does the repository provide a submission form requesting intrinsic metadata in a specific format (to ensure machine readability/interoperability)?
- Does the repository have a long-term preservation plan for the archived data

Data Management Plan – content of the mySNF form

	Question	Help text
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 data sets. (This relates to the *FAIR Data Principles* F2, I3, R1 & R1.2)
	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)
	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)

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)
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)
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 reused. (This relates to the *FAIR Data Principles* I3 & R1.1)

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 are 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 updates, responsibilities, automatic/manual process, security measures, etc.)
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 reuse, 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. (This relates to the *FAIR Data Principles* F2 & R1.3)

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)
	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)
	4.3 I will choose digital repositories that are conform to the FAIR Data Principles. [CHECK BOX]	The SNSF requires that repositories used for data sharing are conform to the FAIR Data Principles. For more information, please refer to the SNSF's explanation of the *FAIR Data Principles*.
	4.4 I will choose digital repositories maintained by a non-profit organisation. [RADIO BUTTON yes/no] → If the answer is no: "Explain why you cannot share your data on a non-commercial digital repository."	The SNSF supports the use of non-commercial repositories for data sharing. Costs related to data upload are only covered for non-commercial repositories.

Examples of data repositories

The following established repositories fulfil SNSF's requirements (allows publishing of FAIR data, non-commercial), and accept datasets from different research fields. It is, of course, possible to archive data on other (field specific) repositories as long as they are in line with the SNSF requirements (non-commercial, FAIR Data Principles)

repository	Non-commercial	Usage of globally unique and persistent identifiers	Possibility to upload intrinsic and submitter-specified metadata	License for data is clearly defined, or can be chosen by submitter	Metadata are always publicly available (also in case of restricted access)	Is the submitted (meta)data machine readable/interoperable?	There is a long term preservation plan for the archived data
Dryad	Yes, all supporting institutions are non-profit (see tab 'institutions')	Yes, uses DOI as permanent identifier system (see tab 'standards')	Yes, basic intrinsic data has to be entered in specific (required) fields. "Free text" fields are available in this same online form to enter additional information. Furthermore, the submitter could upload README files with even more detailed information.	Yes, all data submitted to Dryad is released to the public domain under CC0	On Dryad all data is publicly available under CC0. Dryad only accepts human subject data that is properly anonymized and prepared under applicable legal and ethical guidelines.	Yes, intrinsic metadata has to be entered in a structured way (online form) in required fields. The completeness and correctness of the metadata (e.g. information about the associated publication, the date on which any embargo is to be lifted, indexing keywords) are checked and the DOI is officially registered.	Yes, detailed information is provided

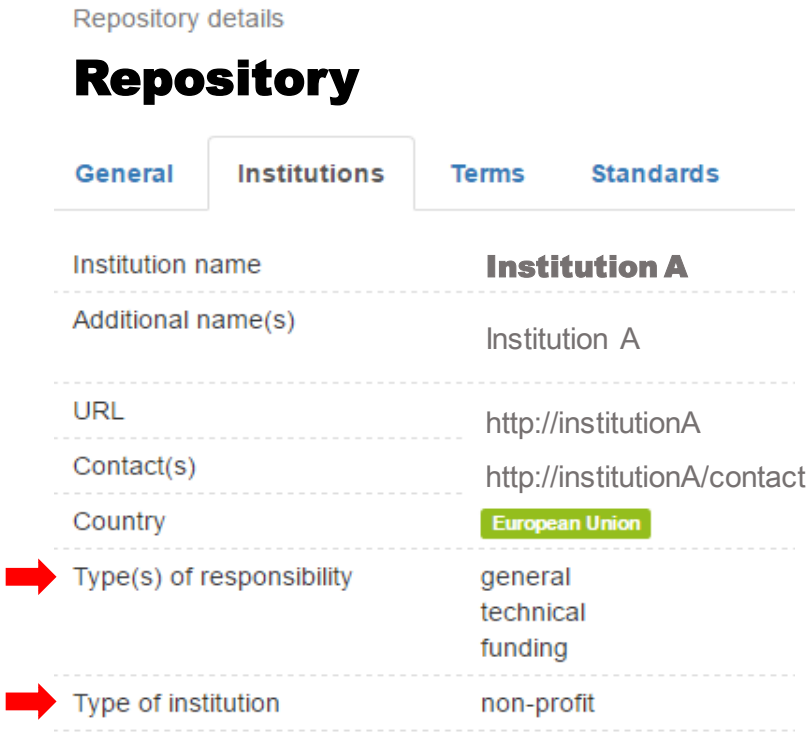
repository	Non-commercial	Usage of globally unique and persistent identifiers	Possibility to upload intrinsic and submitter-specified metadata	License for data is clearly defined, or can be chosen by submitter	Metadata are always publicly available (also in case of restricted access)	Is the submitted (meta)data machine readable/interoperable?	There is a long term preservation plan for the archived data
EUDAT	Yes, all supporting institutions are non-profit (see tab 'institutions')	Yes, uses hdl and DOI as permanent identifier systems (see tab 'standards')	Yes. Metadata is defined in metadata schemas and includes default mandatory fields as well as fields defined by the community under which the deposit is made. EUDAT supports the concept of communities administering their own metadata schemas and publication requirements.	Yes, licenses for datasets are chosen by the submitters and defined via an online form.	Yes, see an example. EUDAT encourages open access to the data, so the default setting for "Open Access" is True, which makes the files publicly accessible by everyone. Switching "Open Access" to "False" will restrict file access to the record's owner and the community administrator. Metadata will always be publicly available.	Yes, basic intrinsic metadata has to be entered in a structured way (online form) in required fields. Metadata is defined in metadata schemas and includes default mandatory fields as well as fields defined by the community under which the deposit is made.	Yes. The service utilises other EUDAT services for reliability and data retention, while storing the data at trusted repositories with national backing, in order to provide a professionally managed and supported IT environment. B2SHARE service providers have agreed on a Memorandum of Understanding with the EUDAT consortium and keep the data accessible for at least 2 years. In the unlikely event that the service will be terminated, the service provider has the obligation to announce this at least one month in advance, and all deposited data and metadata will either be kept accessible or handed over to another EUDAT service provider so that the service can be continued.
Harvard Dataverse	Yes, supporting institution is non-profit (see tab 'institutions')	Yes, uses DOI as permanent identifier system (see tab 'standards')	Yes, basic intrinsic data has to be entered in specific (required) fields. "Free text" fields are available in this same online form to enter additional information. Furthermore, the submitter could upload README files with even more detailed information.	Yes: default is a CC0 waiver, but custom terms of use can be specified.	Yes, see an example	Yes, basic intrinsic metadata has to be entered in a structured way (online form) in required fields. Checks are performed on some fields to ensure that the proper format is used (e.g. email address, date,…)	Yes

repository	Non-commercial	Usage of globally unique and persistent identifiers	Possibility to upload intrinsic and submitter-specified metadata	License for data is clearly defined, or can be chosen by submitter	Metadata are always publicly available (also in case of restricted access)	Is the submitted (meta)data machine readable/interoperable?	There is a long term preservation plan for the archived data
Zenodo	Yes, all supporting institutions are non-profit (see tab 'institutions')	Yes, uses DOI as permanent identifier system (see tab 'standards')	Yes, basic intrinsic data has to be entered in specific (required) fields. "Free text" fields are available in this same online form to enter additional information. Furthermore, the submitter could upload README files with even more detailed information.	Yes: default is CC BY, but user can choose different CC license or restrict the access and set a condition.	Yes, see an example	Yes, basic intrinsic metadata has to be entered in a structured way (online form) in required fields. Checks are performed on some fields to ensure that the proper format is used (e.g. email address, date,...)	Yes

FNSNF
SWISS NATIONAL SCIENCE FOUNDATION

Identification of non-commercial repositories on www.re3data.org

Example of an entry in re3data.org



What qualifies as "non-commercial"?

Type of institution	Type(s) of responsibility		
non-profit	general	✓	non-commercial
	technical	✓	non-commercial
	funding	✓	non-commercial
	sponsoring	✓	non-commercial
commercial	general	x	commercial
	technical	x	commercial
	funding	✓	non-commercial
	sponsoring	✓	non-commercial

Repositories can have multiple institutions. In order for a repository to be considered non-commercial all institutions have to be non-commercial.