FORS GUIDES
to survey methods
and data management

How to draft a DMP from the perspective of the social sciences, using the SNSF template

Diaz Pablo[1] and Stam Alexandra[2]

[1] FORS-UNIL

[2] FORS

FORS Guide No. 07, Version 1.1 (last update January 2025)

June 2019

Abstract:

This guide helps drafting a data management plan (DMP) from the perspective of the social sciences. Drawing on the model developed by the Swiss National Science Foundation, it provides recommendations and tips on how to address the main topics of the DMP so that it fulfils the funder's requirements while at the same time producing high quality data.

Keywords: data management, open data, best practices

How to cite:

Diaz, P., Stam, A. (2019). How to draft a DMP from the perspective of the social sciences, using the SNSF template. *FORS Guides*, 7, Version 1.1, 1-17. https://doi.org/10.24449/FG-2019-00007

The FORS Guides to survey methods and data management

The FORS Guides offer support to researchers and students in the social sciences who intend to collect data, as well as to teachers at university level who want to teach their students the basics of survey methods and data management. Written by experts from inside and outside of FORS, the FORS Guides are descriptive papers that summarise practical knowledge concerning survey methods and data management. They give a general overview without claiming to be exhaustive. Considering the Swiss context, the FORS Guides can be especially helpful for researchers working in Switzerland or with Swiss data.

Editor:

FORS, Géopolis, CH-1015 Lausanne
www.forscenter.ch/publications/fors-guides
Contact: info@forscenter.ch

1. INTRODUCTION

In the context of Open Science more and more funders and ethical boards require researchers to develop a data management plan (DMP). These plans lay out a strategy with respect to how data will be produced/collected, processed and handled throughout and beyond the research project. They encourage the identification and planning of key data management practices that will need to be implemented during the project. Drawing on values such as research transparency (Bishop, 2009) and reproducibility, but also the belief that data often retain unexploited potential (DuBois, Strait, & Walsh, 2018), DMPs aim to ensure that all sharable data will be made available to the research community sooner or later, provided that there are no legal, ethical, copyright, or other issues that would prevent it. Since October 2017, the Swiss National Science Foundation (SNSF) requires that all funding requests include a DMP, and expects funded researchers to make their data available at the time of publication of the respective output.

While a DMP may at first seem like an administrative formality, we contend that it has the potential to be a very useful reflective and practical tool that can help to improve the quality of data and, by extension, research projects. Far from merely serving the ultimate goal of data sharing, careful data management planning is a pre-requisite for high quality research, by ensuring ethical and legal compliance and by making sure that data are collected and described in a way that reinforces their analytical potential.

This guide, which draws on the DMP model developed by the SNSF, provides help specifically for the social sciences. As such, it translates the questions listed in the DMP, geared towards all scientific disciplines alike, into the language of the social sciences, explaining why to do so and giving concrete recommendations on how to address the various questions.

We deliberately decided not to provide 'ready-made' answers to copy and paste into your DMP, since such answers do not exist and may undermine the true value of such a tool. To be effective, a DMP must be customised to the needs and subtleties of each research project. It is about reflecting carefully on the specificities of one's research and addressing a number of important questions that will directly influence the quality of your data and their existence beyond the life of your project.

2. THE DMP TEMPLATE OF THE SNSF

The DMP template as provided by the SNSF[1] is made up of four sections and raises a number of questions, which are listed in the figure below (Table 1). The following will consider each of these questions, briefly explaining why they are important, and then providing practical recommendations on how to address them.

[1] http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf, retrieved June 6, 2019.

Table 1: The four sections of the SNSF DMP template

1. Data collection and documentation ▪ What data will you collect, observe, generate or re-use? ▪ How will the data be collected, observed or generated? ▪ What documentation and metadata will you provide with the data?
2. Ethics, legal and security issues ▪ How will ethical issues be addressed and handled? ▪ How will data access and security be managed? ▪ How will you handle copyright and intellectual Property Rights issues?
3. Data storage and preservation ▪ How will your data be stored and backed-up during the research? ▪ What is your data preservation plan?
4. Data sharing and reuse ▪ How and where will the data be shared? ▪ Are there any necessary limitations to protect sensitive data?

2.1 DATA COLLECTION AND DOCUMENTATION

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

To properly plan the management of your data, it is essential to understand the nature of what you are going to have in your files. Indeed, data management strategies vary greatly depending on the type of data involved. The legal status of the data determines, for example, the degree to which they can be shared and, by extension, the conditions that must be met in order to archive and disseminate them. A clear understanding of your data is also important to properly budget data management costs in your funding application[2].

In this part of the DMP, the main purpose is to provide a description of the data that will be produced/collected/used/reused in your research project. By data we mean: "a reinterpretable representation of information such as numbers, words, measurements, observations or even just descriptions of things in a formalised manner suitable for communication, interpretation, management, decision-making or processing"[3].

There are several ways to define or classify your data. You can do this according to: a) their type; b) the degree of intervention of the researcher in their production; c) their level of processing; d) their format; and e) their legal status. Below, we will briefly present these different typologies in order to offer suggestions for precisely describing your data. We draw your attention to the fact that these different classifications are an aid and not a rigid framework. It is a matter of choosing (or not) what you think is most relevant for defining your data. In addition, it is important to keep in mind that a research project can simultaneously produce many kinds of data.

[2] The SNSF supports the preparation of data and their online storage with up to 10,000 Swiss francs, on the condition that non-commercial service providers are used.
[3] Definition adapted from http://www.dcc.ac.uk/news/representation-information-what-it-and-why-it-important

A) Types of data

There are three main categories of data that research projects produce/collect/use:

- research data
- background data
- process data.

Research data represent the hard core of research. They can be defined as: “recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings”. Depending on the approach and methodology, these may include a wide range of primary and secondary data, such as statistical databases, survey data, interview transcripts, field notes, text corpora, administrative data, social media data, photographs, videos, maps and plans, geolocation data, etc.

Background data (or documentation) consist of important layers of information that enable the understanding and interpretation of research data. It may be created naturally as part of the research process (variable names and labels, syntaxes), or expressly for oneself and others to be able to interpret data (codebooks, questionnaires, metadata[4], or specific documentation like for example on interview contexts).

Process data consist of backstage information that relate to the research process and collection of data. These may include minutes of meetings, email exchanges between team members, contracts, participants’ contact details, consent forms, etc.

B) Degree of intervention of the researcher in the production of the data

Research data can be distinguished by their conditions of production and, more precisely, by the degree of intervention of the researcher in the process. Schematically, research data can be either “naturalistic” or “produced”. Naturalistic data are “data that make up records of human activities that are neither elicited by nor affected by the actions of social researchers”[5]. This can be diaries, minutes, official reports, press articles, etc. Produced data refer to the data generated with the intervention of a researcher. This can be surveys, interviews, observations, etc.

Produced data can themselves be distinguished according to their production conditions. It is customary to distinguish (André, 2014):

- observational data: produced in real time, at a specific moment and therefore not reproducible;
- experimental data: generated in the laboratory, in a controlled context and therefore reproducible;
- simulation data: generated by computer models;
- derived data: resulting from the processing, selection, compilation or aggregation of raw data; and
- reference data: previously published data that are collected, sorted and aggregated.

[4] A definition of metadata can be found on page 8.
[5] http://sk.sagepub.com/reference/download/research/n279.pdf, retrieved June 7, 2019.

Finally, produced data can be either “primary” or “secondary”. Primary data refers to “first hand” data produced by the researcher himself and/or his team as part of a specific research project. Secondary data refers to data produced in another spatial and/or temporal context than the one in which it is used.

C) Level of data processing

Data can be distinguished according to their degree of processing. In particular, it is possible to differentiate raw data from processed data. Raw data refers to the information elements as acquired during the research process: audio or video recordings, observation notes, completed questionnaires, etc. Processed data refer to data that have undergone a transformation since their production. These may include transcripts, aggregated or recoded data, tables, etc.

D) Format

Data can also be distinguished according to their format, that is, the physical and/or numerical structure given to information. This may include text or numbers, audio or video recordings, etc., all in physical or digital form. For example, a research project can generate tape-recorded interviews, computer transcriptions, scans of personal documents, survey data, statistical databases, etc.

E) Legal status

Legally speaking, data can be public, personal, sensitive, or protected by copyright. Public data is information with no existing local, national or international legal restrictions on access or usage that can be freely used, reused, and redistributed by anyone.[6] In Switzerland, the Federal Data Protection Act (DPA)[7] legally defines the central concepts of “personal data” and “sensitive data” (Art. 4). Personal data are defined as all information that relate to an identified or identifiable person.

Sensitive data are personal data that relate to:

- “religious, philosophical, political or trade union opinions, activities and ethnic origin;
- the intimate sphere of the person, in particular, his psychological, mental or physical state;
- individual measures and aid resulting from social legislation;
- criminal and administrative proceedings or sanctions”.

Alone or in combination, these typologies can be useful for giving a more or less precise description of the materials you plan to produce, collect, use, or share as part of your research project.

How will the data be collected, observed or generated?

While the methods of data production may appear to be obvious to the researcher, it is still useful to explain them in order to better define the limits and opportunities in terms of management. The possibility or not of avoiding the unnecessary collection of personal or sensitive data depends, for example, largely on the types of methods used. If in the case of a

[6] https://searchcio.techtarget.com/definition/public-data
[7] https://www.admin.ch/opc/en/classified-compilation/19920153/index.html

survey it is relatively easy not to collect personal data, it is significantly less so when processing audiovisual materials such as video clips.

In this part of the DMP, the aim is to provide a description of the methods and instruments of data collection used. It is also a question of clarifying the strategies for organising these data.

A) Instruments / methods / quality control

The methods of data production / collection can be multiple and varied. It is important here to be as precise as possible and not to limit oneself to general descriptions, such as “qualitative and/or quantitative data”. For example, it is useful to distinguish between:

- semi-directive, non-directive interviews
- participant and non-participant observation
- surveys
- constitution of corpus of documentary materials
- data mining
- laboratory experiments, etc.

In addition to the methods used, it is also important to specify the instruments used to implement them:

- audio and/or video recording
- maintaining a notebook
- use of a questionnaire software
- use of data scraping software, etc.

Finally, it is also necessary to present the way in which the quality will be measured and documented throughout the research. Quality control applies to all the stages of the research, with specific actions directly relating to the methods of data production. The UK data service provides a list of quality assurance actions[8].

B) Organisation

Here it is a matter of demonstrating that a data organisation strategy has been determined (without going into details, however). This can be done through reference to existing guidelines (see below).

Concretely, the folder structure should reflect the nature of the project while taking into account specific organisational needs. File labels should include just enough information, but no more, to find the file that you are looking for. File names must be clear, precise and complete, as well as provide information on the version of the file. As the UK data service puts it, “a version control strategy depends on whether files are used by single or multiple users, in one or multiple locations, and whether versions across users or locations need to

[8] https://www.ukdataservice.ac.uk/manage-data/format/quality

be synchronised."[9] A useful source of information in this context are the recommendations by the Consortium of European Social Science Data Archives (CESSDA)[10].

What documentation and metadata will you provide with the data?

Documentation is a crucial element in the sense that it allows you to understand the data in the context of their production, and therefore to make them meaningful for yourself and others in the future. More precisely, documentation allows for a reflection on the possible (re)uses of the data. Depending on what is considered possible in terms of reuse, more or less documentation may be provided. The possible uses of the secondary data include reproduction and reanalysis as well as meta-analysis of the research project itself (from a history of science perspective).

Here, you are asked to present the types of background data that will be produced. In this respect, it is useful to distinguish between documentation and metadata. Documentation includes any kind of information that is needed for you or others to understand your data in the long-run. It may include information on the research project, the collection tool, fieldwork, data entry and processing, etc. It may also take different forms, such as for example read-me files, codebooks, spreadsheets, statistical syntaxes.

Metadata are generally defined as "data providing information about one or more aspects of the data; it is used to summarise basic information about data which can make tracking and working with specific data easier".[11] Metadata are a form of documentation, structured according to various disciplinary standards. In medical sciences for instance, metadata are usually standards and controlled vocabularies used to document the research process and outputs. They play an important role during the research process. In the social sciences, metadata usually come into play when depositing your data at an archive and therefore at the end of the research process. Archives and other institutional repositories require you to complete metadata fields that follow international standards and that will ultimately contribute to making your data findable by humans and machines alike (e.g., keywords, methodology, author's name, etc.).

2.2 ETHICS, LEGAL AND SECURITY MEASURES

The aim of this section of the DMP is to specify the ethical and legal issues raised by the research project and to present the measures planned to ensure: 1) the protection of participants, 2) security of the data, and 3) respect of copyright.

How will ethical issues be addressed and handled?

Ethical aspects should be at the heart of the reflections of social science researchers. The interest of this section of the DMP is that it encourages researchers to be aware of the ethical implications of their research projects and to put the interests of participants at the centre of the reflexion. Far from being confined to a "set of values and principles", ethics is "a reasoned reflection with a view to doing the right thing". Doing research in an ethical way therefore implies finding a balance between what research demands, the duty of protection, and institutional requirements.

[9] https://www.ukdataservice.ac.uk/manage-data/format/versioning.aspx, retrieved June 6, 2019.
[10] https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
[11] http://www.seismofaults.eu/index.php/component/seoglossary/2-technical-terms/metadata?Itemid=152

This section should contain information on the researcher's strategy to ensure that the rights and integrity of participants are respected during and after the project. More specifically, it must indicate:

- What are the risks (physical and/or psychological) faced by participants when taking part in this study? Are the expected benefits of the study greater than the risks involved?
- What standards apply to the types of data generated by your project? For example, if you plan to work with sensitive data, you must show that you are aware of the legal requirement to obtain the explicit consent of the participants. If your research project falls under the Swiss Human Research Act (HRA), you must mention your obligation to submit it to the cantonal ethics commission for evaluation. Be aware also that some universities make it mandatory for research projects to be submitted for ethics review to their internal committees.
- What measures are planned to obtain the necessary authorisations for the collection, processing and sharing of personal/sensitive data? Here it is a question of describing and justifying how participants will be informed and how consent will be obtained in practice. You can, for example, give information on: the amount of information that will be given to the participants (subject of the study, aims, hypotheses, etc.); the way in which the information will be given to the participants (oral, written, at the beginning of the project, at the end of the project, etc.); the scope of the consent sought (free use of data, use of anonymised data, authorisation to archive/share data, etc.); the manner in which consent will be obtained (oral question, written form, etc.). For more information about informed consent, see our FORS Guide on consent (Kruegel, 2019).
- If the research provides for only partial information to the participants or even deception, this must be (scientifically) justified.
- What are the measures planned to protect personal or sensitive data, once produced? You can, for example, specify whether you intend to use pseudonymisation and/or anonymisation techniques. The justification of your choice must be based on the sensitivity of the subject, the vulnerability of the population studied, and the risks involved.

For more detailed information on the ethical issues of social science research, see our FORS Guide on ethics (Diaz, 2019).

How will data access and security be managed?

Data security is a central issue in data management. Indeed, the Federal Data Protection Act requires that personal data “be protected against unauthorised processing through adequate technical and organisational measures” (art. 7). However, security is often not well controlled. For example, few researchers are aware of the risks associated with using platforms such as Dropbox, Google Drive, or iCloud, to store and transmit their sensitive data. Indeed, most of the solutions available at the international level are owned by companies (generally based in the USA) whose operations provide for the extraction and use of data for commercial purposes. Using this type of tool to store personal and/or sensitive data therefore presents a significant risk of violating the privacy of research participants.

In this section, you are asked to describe the measures planned to prevent any illegitimate access to the data (e.g. hacking, theft, etc.) during production, processing and storage as well as those provided to prevent any damage or loss. In other words, it is a matter of presenting the technical, organisational, physical and digital procedures set up to secure research materials.

Regarding access, it is possible, for example, to describe solutions such as: avoiding the use of data collection and processing devices that allow data to be extracted by their suppliers for commercial use (Smartphones, online survey softwares, online transcription tools, online PDF mergers, etc.); setting up passwords on all the computers used; encrypting hard disks; using institutional servers; adopting a clear policy for granting computer access; prohibiting the team from using cloud-based solutions; and using computers not connected to the Internet for the analysis of certain materials. Regarding loss or damage, the measures described may include regular data backup, the use of multiple (secure) storage locations, the use of institutional servers, the use of durable physical and digital formats, or the prohibition for the research team to store data on unreliable media such as USB flash drives.

If you decide to use institutional storage solutions, you can also ensure that they are properly certified (see for example ISO27001, CoreTrustSeal, FedRamp). For more information, see the CESSDA recommendations on security..[12]

How will you handle copyright and intellectual property right issues?

At the beginning of a research project, it is essential to agree on the ownership of the data that will be generated. At a time when research projects are increasingly collective and international, this issue is becoming particularly important. For example, it is necessary to clarify early on who owns the data used in doctoral theses carried out within the framework of a funded project. Are they the property of the principal investigator or doctoral student? The same applies to data generated by national partners in international research.

In this section you are asked to identify who has which rights to the data. In Switzerland, for example, most universities own the databases produced by their employees, although the latter retain their intellectual property rights. Funders such as the SNSF do not generally claim intellectual property rights. If an interview is filmed (and not anonymised), the interviewee may also have a number of rights (image rights, intellectual property, etc.). Finally, in the case of multi-centre research, agreements can be signed that grant rights to all parties (e.g. data sharing agreements or research agreements). The rights of each other over the materials used must therefore be clarified according to the institutional and legal framework as well as the type of materials used. Videos, for example, do not pose the same type of challenges as literary works.

This section should also contain some information about the licenses that will be applied to the data for re-use (restrictions, etc.). The most used types of licenses today are Creative Commons. For a discussion on why and how to license data, see Ball (2014).

[12] https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/4.-Store.

2.3 DATA STORAGE AND PRESERVATION

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

The importance of this question is obvious: failure to properly store and back-up your data may result in data loss or inappropriate access. Not only are you at risk of losing valuable data, but the consequences can potentially be dramatic for your respondents, should sensitive/personal data not be anonymised or properly protected.

In this section you are expected to describe where your data will physically be stored during the research process and how you will protect them from loss. A first step is to study the existing solutions that are available at your institution with respect to data storage and security measures. Most academic institutions offer secure storage solutions with data usually being stored locally on an institutional server. We recommend you always start by considering institutional solutions, and assess whether they meet your research needs as well as the requirements of ethical boards. If they do meet your needs, then we recommend you use them as much as possible to store your data and files. Such servers are usually much safer than your own computer hard drive, which can easily be hacked, damaged or lost. They usually provide controlled access, and back-ups are carried out on a regular basis. Furthermore, it is often possible to access them remotely with your institutional credentials. It is of course possible to use (in parallel) your own computer, but be aware that you are at higher risk of data loss. If you chose to work with your personal computer, it is important that you protect any personal and sensitive data you may have by encrypting the data in a password-protected folder. Also, in case your data are not on an institutional server, make sure you regularly make copies of your data on different devices (hard drive, secure cloud solutions, etc.). Some software solutions have been developed to back up your data on a cloud (i.e. CrashPlan).

If your institutional repository does not meet your project needs, then you need to work out an alternative. You may for example be part of a larger team with colleagues located in different national and institutional institutions. In such cases, it may not be possible to create a shared folder on your institutional server. You may want to consider using cloud solutions, or generate your own server (using Nextcloud for example). Be very careful with cloud solutions and make sure they are located in your country (e.g. SWITCHdrive in Switzerland). That said, sensitive and personal data should always be encrypted.

While filling in your DMP, you may also consider other security issues, if relevant to your research project, such as data transfer. For example, you may need to transfer raw data to someone you hired to transcribe the data and who cannot work onsite. E-mails are unsecure and should never hold sensitive information. Best is to encrypt data (e.g. with Veracrypt) and either hand them over in person (using for example a portable device) or send them through a local cloud storage (e.g. SWITCH file sender). If the person works in the same institution, you may be able to create a shared folder. In the case of transferring sensitive data, make sure the receiver has full rights to access them (i.e. consent) or has signed a data protection form.

In short, security solutions depend on your research needs. Always start considering institutional options, and if these are not suitable, start looking for other solutions. In your DMP you need to show you have a clear understanding of the various security issues you may be facing. Whenever possible list the solutions, or define actions to overcome particular

issues. For a more developed overview and further recommendations, consult the CESSDA online module on storage and back-up.

What is your data preservation plan?

It is important to anticipate future possible uses of your data to make sure you preserve important data, or data with further potential, be it research data, background data or process data. We tend to keep everything, including many versions of the same file 'just in case', which not only eats up lots of space, but also affects data retrieval and readability. The more we keep, the more maintenance work is required to make sure formats are upgraded alongside technological changes.

In this section of the DMP you are expected to explain how you will select the data you want to keep in the long-run, and make sure they remain accessible over time by storing them on an adequate device and choosing an appropriate preservation format. Data preservation applies to any data you have collected, including research data, background or process data.

Both, devices and data format will need to be updated over time. Note that most data archives take charge of the preservation of any data you deposit with them (see section 4 below). You do not need to provide a detailed preservation plan at this point, but rather show that you have considered a number of questions and are aware of the importance of ensuring long-term preservation of at least part of your data.

If you are unsure about what data to keep in the long run, do not hesitate to contact your data archive/repository for advice.

A) Data selection

Data selection applies to data you will ultimately share as well as the data you do not intend to share but wish to keep for yourself. By the end of a project we have often accumulated a large amount of data and documentation, including many versions of the same data. You therefore need to ask yourself the following questions:

- What data do I want to keep beyond the research project? What are the most likely purposes? Consider all your data (research data, background data, process data) and their various formats. You may for example decide to keep interview transcripts but delete recordings.
- How long do I want to keep the data and why? You may for example anticipate future research, or keep or delete some (sensitive) data for legal reasons.

B) Long-term preservation

Long-term preservation implies making sure data remain accessible over time, both in terms of their physical location and their readability over time. Ask yourself the following questions:

- Will any of my data be deposited in a data archive or repository?
- How will I ensure data preservation? Where will I store my data, and how will I make sure my data will remain accessible over time? This involves both updating file formats and making sure the devices on which they are stored are not obsolete.

In general, we recommend that you favour open source formats as opposed to proprietary formats. To this end, the Swiss Federal Archives have produced a useful guide on archivable file formats[13]. Be aware that some formats cannot be updated into an archivable format at the end stage. This is the case for relational databases such as Filemaker for the humanities. If possible, favour an open source solution (for example SALSAH) from the very start of your research project.

2.4 DATA SHARING AND REUSE

How and where will the data be shared?

It is important to think about the future of your data at an early stage, to make sure you collect and prepare your data in a way that allows its use, but also so that you can benefit during the project from the guidance of the repository that will host your data.

Most repositories provide persistent identifiers (such as Digital Object Identifiers- DOIs) that allow your dataset to be cited and accessed over time. Furthermore, data deposited within established repositories are curated according to standards that follow international and/or disciplinary standards, such as the Data Documentation Initiative (DDI) in the social sciences.

To answer this question, you need to have a clear view about what data you will likely make available with the corresponding scientific outputs during the course of the research and possibly after the project ends (in case you choose to make more data available for re-use purposes). The point here is to describe your intentions and review them as the project develops. Ideally, you will also identify a repository where you will be depositing your data. It is worth it to contact the repository at an early stage, to make sure you prepare the data in a way that will help you save time at the end of your project. Further, repositories may also provide useful guidance on day-to-day data management, which will help you implement your data management planning strategy.

Data can be shared through different channels, the most common being:

- institutional archives or repositories
- trusted disciplinary archives or repositories
- general data sharing services (figshare, Zenodo, Dataverse)
- journals
- by the researchers themselves (personal website, data transfer)

As a rule, we recommend that you favour institutional archives or trusted disciplinary archives, since these are the most likely to provide an infrastructure that meets funder's requirements as well as to enable long-term preservation. Check out whether your university offers a satisfactory solution or whether a national solution exists, preferably discipline-specific. Please note that many repositories are more like storage zones and do not ensure preservation, unlike data archives.

[13] https://www.bar.admin.ch/bar/fr/home/archivage/versement-de-documents/documents-numeriques.html.

It is therefore important that you make sure the repository or archive follows the FAIR principles[14] in accordance with the SNSF requirements, allowing data to be Findable, Accessible, Interoperable and Reusable. A list of international repositories can be found on the following website: https://www.re3data.org/[15].

FORS provides an online platform, called SWISSUbase, which allows you to access social science data as well as archive your own research data, be it quantitative or qualitative. SWISSUbase is aligned with the FAIR principles and is therefore eligible as a repository for SNSF projects. We accept all social science data from researchers based in Switzerland, as well as data of interest for the social sciences. To access data end-users must be registered in our system and sign a user contract each time they request data. As a researcher who deposits data within ur system, you will be able to set up access parameters, from 'open to all registered users' to restricted access in accordance with the funder's policy. More information on SWISSUbase can be found online (https://www.swissubase.ch). Should you wish to deposit your data with FORS, or for any other question, do not hesitate to contact us at dataservice@fors.unil.ch.

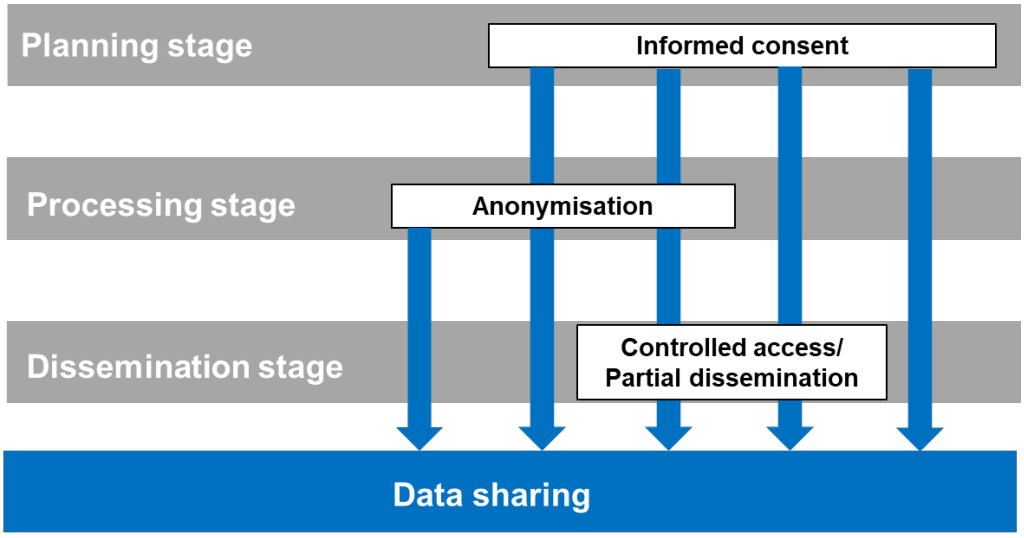
Are there necessary limitations to protect sensitive data?

By answering this question you can determine different layers of protection that will allow the sharing of your data while meeting legal and ethical requirements. The collection of sensitive data is not in itself a reason for not sharing data. It is important to figure out the various options and decide what solutions can be put in place to allow safe sharing of all or part of your data. Some solutions need to be implemented at an early stage, such as obtaining informed consent, while others can be implemented at a later stage (anonymisation, controlled access). It is therefore important to answer this question to make sure that you do not miss out on some necessary actions, and can start planning for it at an early stage.

To answer this question, you need to consider the nature of your data and identify whether you will collect or process any sensitive data (see question 1: what data will you collect, observe, generate or re-use?). Should you have any personal and/or sensitive data, then consider the following ways to protect them: a) by obtaining consent, b) by anonymising your data, and c) by controlling data access. Sometimes only one layer of protection is sufficient, while other times you may need to combine all three (Figure 2).

[14] http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf, retrieved June 7, 2019.
[15] Beware that not all repositories on the list meet the FAIR principles.

Figure 1: measures to protect personal and or sensitive data



Informed consent[16] is a precondition for sharing sensitive data. Legally in Switzerland, if respondents give their formal approval, you are allowed to make personal and sensitive material publicly available. In most cases however the value of the data does not rely on such information and therefore it is best practice to anonymise or pseudonymise[17] your data. Should your data be fully anonymised, then you do not need, in theory, any informed consent. Since this is very difficult to guarantee, especially when it comes to qualitative methodologies, we recommend that you always seek consent for sharing the data, explaining what measures (e.g., pseudonymisation) will be applied.

Sometimes (pseudo-)anonymisation is not possible, since your data would lose much of their value, or might, even if informed consent was received, fail to fully protect your respondents. You may therefore add an extra layer of protection by choosing to place your data within a repository that allows you to control and customise access conditions to your data in accordance with your funder's policy. Depending on the nature of your data and the associated risks, you may set up access conditions, ranging from anyone to selected audiences (e.g. for research purposes or for teaching purposes). Some repositories such as SWISSUbase also allow you to agree on a case-to-case basis for your data to be disseminated. In such cases you will be asked for approval each time someone requests your data. We, recommend, however, that you only resort to this option when it is really needed to protect your data and respondents.

Finally, established archives usually have end-user contracts, which means that in order to be able to access your data, third parties need to sign a contract whereby they promise to follow a certain number of rules, including the obligation not to try to identify respondents.

[16] See our FORS guide on informed consent: https://doi.org/10.24449/FG-2019-00005.
[17] To find out more about anonymisation and pseudo-anonymisation, see the guidelines prepared by the Finish Social Science Data Archive: https://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html.

3. CONCLUSION - RECOMMENDATIONS

Recommendation 1 – Keep in mind that good data management is not only a pre-requisite for data sharing: above all it allows you to conduct high quality research by strengthening your data and their analytic potential.

Recommendation 2 – The cost of managing your data only at the end of your research project is much higher than applying good practices throughout the research process. Data management planning strongly contributes to reducing costs, not only as it allows you to identify key actions and to make sure you implement them at the right moment, but also since it allows you to budget data management costs in your research proposal.

Recommendation 3 – While the Swiss National Science Foundation requires that the data underlying scientific publications must be made available provided that there are no legal, ethical or copyright issues, we strongly recommend that you share any data that are relevant for re-use. Data underlying publications must be made available at the time of the publication, while wider data can be made available after the project ends.

Recommendation 4 – Once your project gets funded, do not forget to review your DMP and further develop your data management strategy. A DMP should be a living document. It should be revised and developed as the research project is carried out.

Recommendation 5 – If you know in which archive/repository your data will be deposited, do get in touch with them at an early stage in the research process. This may help you save time by ensuring you handle the data in a way that will facilitate data sharing, but also by benefiting from their expertise and recommendations with respect to day-to-day data management practices.

4. FURTHER READINGS AND USEFUL WEB LINKS

If you are interested in practical data management tips and recommendations, you might look at the CESSDA data management expert guide:
https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide

If you are interested in standards for archiving digital documents, you may visit the Swiss Federal Archives website: https://www.bar.admin.ch/bar/fr/home/archivage/versement-de-documents/documents-numeriques.html

If you are interested in practical data management tips, take a look at the FORS Guides to Data Management. These guides provide essential support, from planning to sharing data, including data protection and ethical issues, with a focus on the Swiss context: https://forscenter.ch/publications/fors-guides

REFERENCES

André, F. (2014). Chapitre 5. Déluge des données de la recherche: Petit manuel d'immersion. Curation, infrastructures et partage. In L. Calderan, P. Laurent, H. Lowinger

& J. Millet (Eds.), *Big Data: Nouvelles partitions de l'information. Actes du séminaire IST Inria, october 2014* (pp. 77-95). Louvain-la-Neuve, Belgique: De Boeck Supérieur.

Ball, A. (2014). How to license research data. *A Digital Curation Centre and JISC Legal'working level' guide*. Retrieved June 07, 2019 from http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf

Bishop, L. (2009). Ethical Sharing and Reuse of Qualitative Data. *Australian Journal of Social Issues*, 44, 255-272. doi:10.1002/j.1839-4655.2009.tb00145.x

Diaz, P. (2019). Ethics in the era of open research data: some points of reference. *FORS Guides*, 3, Version 1.0, 1-18. doi:10.24449/FG-2019-00003

DuBois, J.M., Strait, M., & Walsh H (2018). Is it time to share qualitative research data? *Qualitative Psychology,* 5(3), 380–393. doi:10.1037/qup0000076

Kruegel, S. (2019). The informed consent as legal and ethical basis of research data production. *FORS Guides*, 5, Version 1.0, 1-14. doi:10.24449/FG-2019-00005

Swiss Federal Archives. (2018). *Standards for archiving digital documents – archivable file formats*. Retrieved June 16, 2019 from https://www.bar.admin.ch/bar/fr/home/archivage/versement-de-documents/documents-numeriques.html