Research Data Management Handbook

Written in collaboration with Worcester Polytechnic Institute
Authored by Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, and Jeffrey Xiao
11 October 2018

WPI

UniversitätsSpital Zürich

Table of Contents

Chapter 1: Introduction 2

- Purpose of this Handbook 2

Chapter 2: Documenting Data 3

- Paper Lab Notebooks 3
- Electronic Note-Taking 3
- Audio Note-Taking 5
- Code 5
- Applications to SNSF Requirements 5
- Conclusions 6

Chapter 3: Naming Data Files 7

- Necessary Components 7
- Examples of File Names 7
- Common Mistakes 8
- Applications to SNSF Requirements 8
- Conclusions 8

Chapter 4: Organizing Data File Folders 9

- Three Characteristics of Effective Organization 9
- Methods 9
- Applications to SNSF Requirements 10
- Conclusions 11

Chapter 5: Storing and Transferring Data Files 12

- Handwritten Data 12
- Electronic Data 12
- Applications to SNSF Requirements 13
- Conclusions 14

Additional References 15

Paper Lab Notebook Checklist 16

Electronic Note-Taking Checklist 17

Audio Note-Taking Checklist 18

Naming Files Checklist 19

Organizing Data Files Checklist 20

Storing and Transferring Physical Data Checklist 21

Chapter 1: Introduction

Purpose of this Handbook

Managing your research data is a highly important task. Considering and planning how data should be created, named, stored, and organized at the onset of an experiment or project will save you both time and effort when trying to analyze and process results at the end. Doing so will also prevent any later confusion and allow you and others to quickly identify and locate information.

The purpose of this document is to provide recommendations for efficiently and effectively documenting and organizing research data. In the following chapters, you will be able to find guidance on:

- Documenting Data
- Naming Data Files
- Organizing Data File Folders
- Storing and Transferring Data Files

Additionally, for projects wishing to apply for grants from the Swiss National Science Foundation (SNSF), specific guidance on how to best manage your data in compliance with FAIR Data Principles and the Data Management Plan (DMP) can be found in each chapter.

Chapter 2: Documenting Data

In this section, best practices for documentation in both electronic and paper mediums are described.

Documentation, as it relates to research data, refers to information that aids in the interpreting and understanding of data. The act of documenting data entails specifying what, when, how, and where an experiment was conducted. Comprehensive documentation not only enables others to understand the work that was conducted but also serves as clear reminders to the person who conducted the work if the data should need revisiting.

Paper Lab Notebooks

The following are good rules of thumb for documenting information in paper lab notebooks, and an example of these practices can be found on the next page:

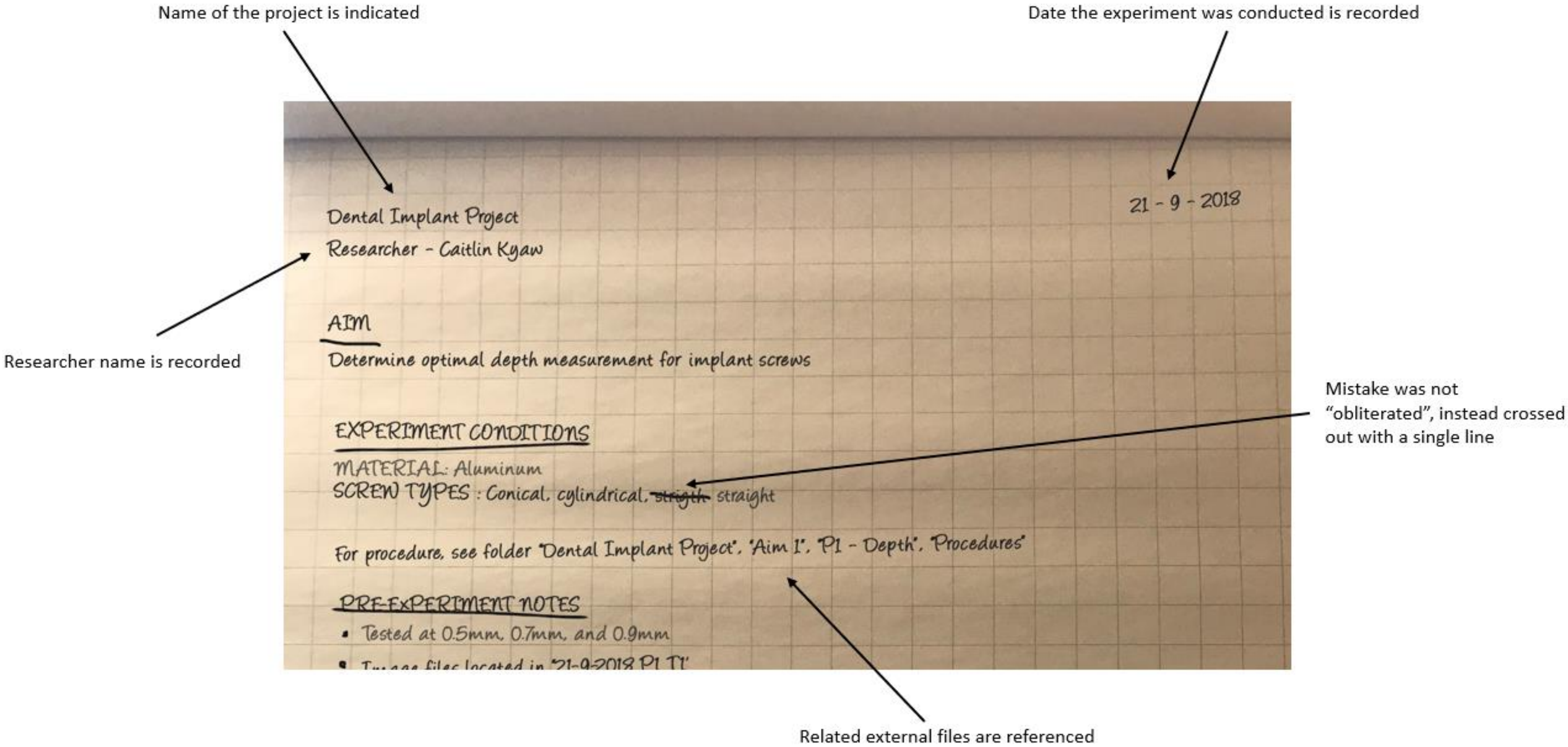
- **Create and update a table of contents.** This allows you to quickly find experiments in lab notebooks that have many entries.
- **Date every page.**
- **Write clearly and neatly.** Legible handwriting benefits yourself and others when trying to read and understand the information later.
- **Record the name and aims of each experiment or project.** This information is ideally located at the top of each page and allows you to remember what the goal of the experiment or project was at a later time.
- **Record the procedure in detail.** Doing allows experiments to be easily reproducible. Be specific when describing details such as techniques, amounts, machines used, and materials.
- **Record experimental conditions in detail.** Similarly to recording the procedure, this is important because it makes experiments easily reproducible. Conditions can include temperature, brand and lot numbers of chemicals, the duration of the experiment, instrument calibration numbers, etc.
- **Record observations and results in detail.** Details of the experiment or project can be clear to you at the moment, but you may forget certain aspects years, months, days, or even hours after completing an experiment. Similarly, things that may be clear to you may not be clear to others. Additionally, if there were mistakes or failures relating to the experimental results, explain why.
- **Do not "obliterate" mistakes**. To correct them, cross them out with a single line, and briefly explain why the crossed-out information should be disregarded.
- **Explain and expand abbreviations.**
- **Attach any external graphs or computer printouts using glue or other permanent binding materials.**

Electronic Note-Taking

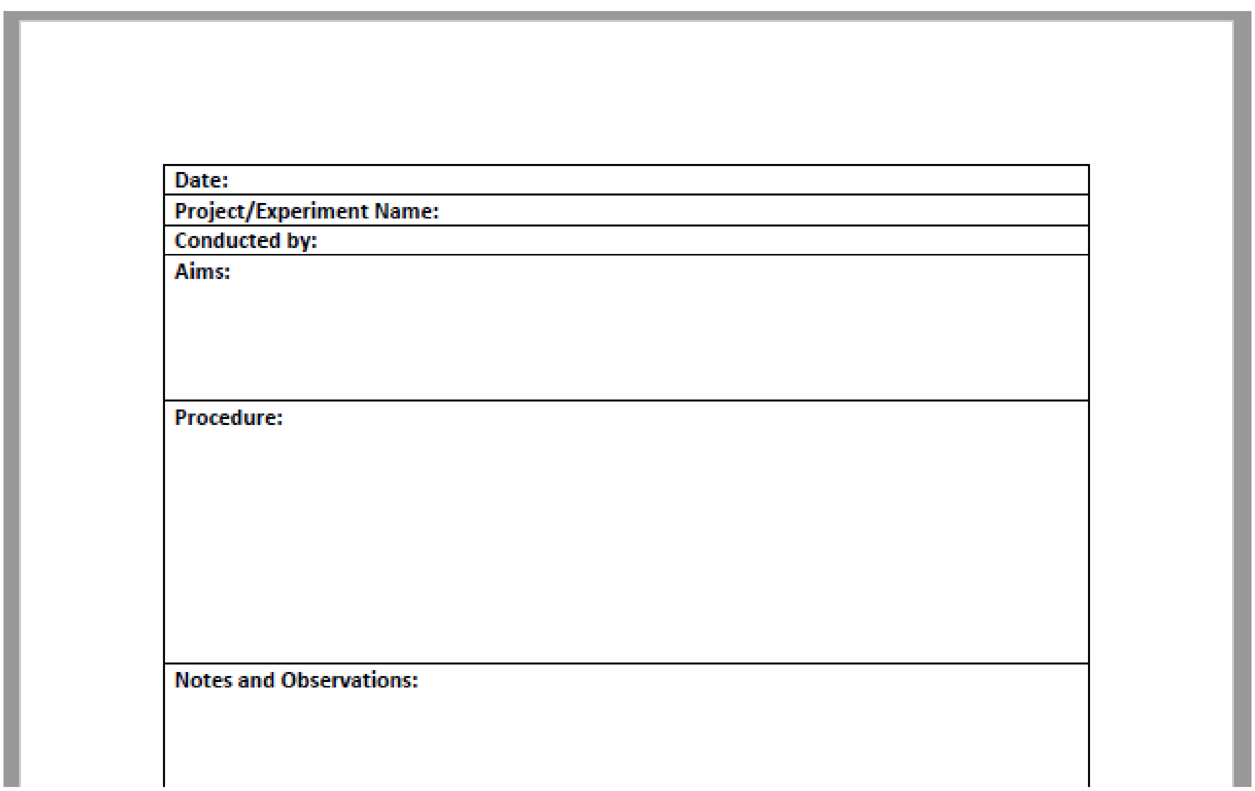
The following are good rules of thumb for documenting information when taking notes electronically:

- **Document the date.**
- **Explain and expand abbreviations.**
- **Record the name and aims of each experiment.**
- **Record experimental conditions, observations, and results in detail.**
- **Import any relevant images and add annotations.** This allows any related observations to be found in one place.
- **Create note-taking templates for common projects and experiments.** Doing so saves you time when repeating common experiments. An example can be found on the next page.
- **Share the above templates with colleagues.** This standardizes the way projects and experiments are documented, allowing for greater consistency.

The following is an example of a well-documented paper lab notebook:



The following is an example of an electronic note file template:



Date:
Project/Experiment Name:
Conducted by:
Aims:
Procedure:
Notes and Observations:

Audio Note-Taking

The following are good rules of thumb for documenting information when taking notes electronically:

- **Speak clearly.** It is a good idea to test if you can be heard clearly on the recording beforehand.
- **State the date.**
- **State who is taking the recording.** If it is just you, state your own name. If several people will be included in the recording, be sure that they also state their names.
- **State the name and aims of each experiment or project.**
- **State experimental conditions, observations, and results in detail.**
- **Verbally relate the recording to any relevant notes (handwritten or electronic) that are taken concurrently.** This allows you to group observations from the same project together.

Code

Clearly and concisely documenting code used for processing and analyzing data is also highly important. The following are good rules of thumb for documenting code:

- **Name variables, methods, and classes clearly and consistently.** This allows viewers to quickly understand how certain items are related and used.
- **Only annotate your code if necessary.** If many comments are needed to clarify your code, you should probably write the code differently. However, comments are acceptable in moderation.
- **Be concise.** Remove any irrelevant lines of code. If any code has been commented out, you probably don't need it.

Applications to SNSF Requirements

Data Management Plan (DMP)

Documentation of research data is highly important when it comes to the submission of the DMP. The first section of the DMP will ask you to explain what data will be collected, and how you intend to collect it, and what metadata, or data that describes data, will be included. To be able to accurately detail such information, consider documenting the following during an experiment or project:

- Data type
- Format of raw data
- Format of processed data
- Estimated volume of raw and processed data

An example of this might look like:

> *Western blot data is collected for the detection of proteins. Images are taken using a CCD camera. The image data is quantified and processed in Microsoft Excel. The data files for the images and Excel files total 2GB and 4GB, respectively.*

To be able to describe the metadata and how it will be collected and provided, consider documenting the following during an experiment or project:

- Content
- Responsibilities
- Software
- Organization

An example of this might look like:

> *Documentation for this data describes experimental conditions such as temperature (Celsius) and humidity, as well as the experimental protocol used. It is the responsibility of the researcher to detail the conditions of the experiment daily. Documentation is in a text file format, and the file is named and organized in the pre-agreed format of "YYYYMMDD_Researcher_ProjectName.txt".*

FAIR Data Principles

Documentation, as it pertains to the FAIR Data Principles, is specifically related to findability and reusability. In order to be findable, the data should be described clearly and thoroughly by metadata. The context, conditions, and characteristics of the data are all important aspects of metadata, and other researchers should be able to easily understand the nature of the data. Similarly, to be reusable, it is your responsibility to ensure that information such as the scope of the data, any limitations of failures of the experiment, the date the experiment or project was conducted, and other additional details are included in your documentation.

Conclusions

Documenting your data thoroughly will allow you and other researchers to quickly understand the aims, conditions, observations, and results from a project or experiment. Although it may take longer to include more details, clear documentation will save time during the processing and publication phases.

Chapter 3: Naming Data Files

This section describes how to name data files in a descriptive and effective manner.

Descriptive file naming is highly important when conducting research. A clear and detailed file name enables you and your collaborators to quickly and easily identify files. Furthermore, having standardized file names across projects, research groups, and departments allows for additional clarity in research.

Necessary Components

The following components should always be included in your file name:

- **Date.** The recommended date format is YYYY_MM_DD. This will allow you to easily sort your file names by date in a file manager or explorer.
- **Researcher Name.** Incorporating your name or your initials into the file name allows collaborators and future researchers to determine who collected the data.
- **Experiment Type.** Listing the experiment type allows you to quickly identify the appropriate data files within each project subfolder.
- **Trial Number.** This allows you to identify exactly which iteration of an experiment the file describes and differentiates it between similar experiments.
- **Keywords.** These can either be included as full words or as abbreviations and describe additional parameters or conditions of an experiment. If you choose to use an abbreviation, be certain that it is used in the same way throughout your research group to avoid any confusion.

In addition to being descriptive, file names should be standardized across projects, research groups, and collaborations. Standardization can be accomplished in one of two ways:

- **Discussion.** Discuss and define how you and your project team would like to name files. Have this decision recorded somewhere and hold each other accountable for using it.
- **File Naming Application.** Utilize the file naming application, File Namer Pro, which automatically generates a standardized file name for you. This is the preferred method.

When file names are standardized, researchers do not have to spend time trying to discern what a files' contents are, and can easily glean information by looking at the file name. This also allows files to be easily located in a file manager. If you need to look for a file created on 10 September 2018, you simply need to look up files with "2018_09_10".

Examples of File Names

The following is an example of a file that has been ***poorly*** named, based on the information provided below:

Date: *10 September 2018*
Researcher Name: *James Bond*
Experiment Type: *Sound Check*
Additional Keywords: *40hz, Stealth Car*
File Name: *SoundCheck1*

This is a poor file name because it does not provide very descriptive information. "SoundCheck1" could refer to any soundcheck experiment done by any person, on any date, with any parameters.

The following is an example of a file that has been named ***well***, based on the information provided below:

Date: *10 September 2018*
Researcher Name: *James Bond*
Experiment Type: *Sound Check*
Additional Keywords: *40hz, Stealth Car*
File Name: *2018_09_10_JB_SC_40hz_StealthCar*

This is a good file name because it specifies who conducted the soundcheck experiment, when, and uses specific and descriptive parameters.

Common Mistakes

- **Making long file names.** While including many keywords is useful, be certain to keep your file name within a reasonable character limit.
- **Using inconsistent abbreviations.** Your abbreviations may make sense at the moment, but might not several weeks or months later. Be certain to create and maintain a glossary of keywords so you can easily decipher your file names.
- **Using special characters.** Avoid using special characters such as “%” and “&”, as these symbols make reading file names difficult and are often not program-friendly.

Applications to SNSF Requirements

Data Management Plan

The Data Management Plan also requires information concerning how files will be named. As stated above, you and any collaborators should reach a consensus on how files will be labeled, and ensure that the agreed-upon convention is consistently used. In your DMP, explain clearly what the convention is. An example of this might look like:

All raw data will be named according to the pre-agreed format of “YYYYMMDD_Researcher_ProjectName_ExperimentName_Condition1_Condition2”. A concrete example of this is “20180903_ER_UJI_DM_7mm_Al”, where the date is 3 September 2018, the experiment was conducted by “Example Researcher”, the project name is “Upper Jaw Implant”, the experiment name is “Depth Measurement”, and the conditions of the experiment are a depth of 7 millimeters and aluminum material. Information detailing any non-standard acronyms can be found in the project text file.

FAIR Data Principles

Standardized, specific file naming allows data to be findable, in accordance with FAIR Data Principles. To be findable, each dataset should be designated with a unique identifier that clearly differentiates it from other datasets. This identifier should be consistent through all mentions and references to the dataset and be descriptive. A good file name will allow you and others to quickly identify a file and its contents.

Conclusions

Naming your files in a detailed, standardized manner will allow you and others to quickly and efficiently discern a file’s contents without needing to open it. This saves time during the processing and publication phases, allowing you to focus on your research instead of looking for files.

Chapter 4: Organizing Data File Folders

In this section, best practices for organizing file structures are discussed.

Maintaining a proper file structure system over the course of an experiment allows you to quickly and efficiently locate and identify research data files. The time that would have been spent conducting a search similar to an archeological dig could be better spent on other research projects when files are properly organized.

Three Characteristics of Effective Organization

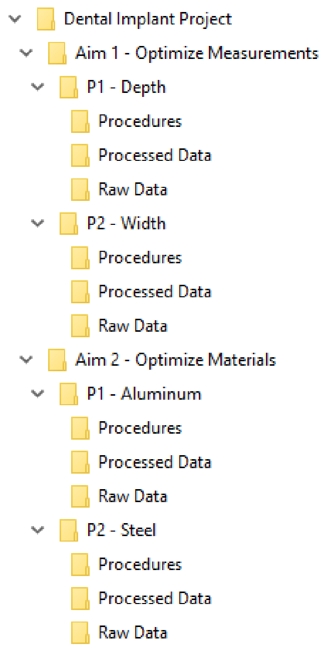
There are several characteristics of a proper file organization system:

- **Data is easily filed.** The file organization system should have a logical file structure rather than a maze of folders and files. This will create an easier process to store files for both the user and the machine.
- **Files are easily found.** The organization structure should allow you to easily locate files via a search function or by eye. If it is confusing to figure out where to store a file, it'll probably be just as confusing to find it later.
- **File folder structures are standardized.** File organization methods should be consistent across projects and departments, only differing slightly depending on the project.

Methods

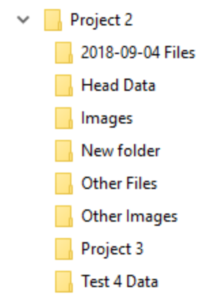
The following are good rules of thumb for organizing data files:

- **Maintain folder hierarchies.** Designating a folder for the project as a whole will differentiate between projects. Within the main project folder, subdirectories can be created. An example hierarchy can be: Project -> Experiment -> Sample. In this way, general information about the project and data files from experiments will cause confusion.
- **Create folder structures for stages, tasks, or aims of the experiment.** Creating subfolders within the main experiment folder for specific stages, aims, or tasks of an experiment or project can help track the progress of an experiment. Examples of folders in this structure could be Planning, Data Collection, Results, Documentation, Future, Objective 1.
- **Group folders by date.** A common and important experiment parameter is the date. Grouping projects or experimental data by date can simplify the process of searching for the most up-to-date files.
- **Minimize the number of folders.** Having a large number of folders can make it overwhelming and difficult to locate files in the future. Have fewer folders and more descriptive file names to easily locate files.
- **Limit the number of files and folders on the desktop.** Although storing files and folders on the desktop can seem like an easy way access them at the time, it becomes difficult to track as more and more files and folders are added.
- **Name files and folders logically**. Refer to the Naming Files chapter for more information.
- **Discuss folder structure standards with collaborators.** The whole group should come to a consensus on organizing data that can consistently be followed on a day-to-day basis with minimal changes. In addition, these templates and standards should be reused for other projects to enforce that folders be named a specific way each time.
- **Create shortcuts to frequently accessed folders.** If you find yourself returning to a certain folder often, it is a good idea to make a shortcut in file explorer or at the bottom of the menu bar. This way, you eliminate the need to shuffle through folders every time.



The folder structure shown on the left is an example of a well-organized file structure. The project itself has a main folder, and subfolders are in a logical hierarchy. Additionally, the folders are named clearly.

The folder structure show on the right is an example of poorly-organized file structure. There is no logical hierarchy of folders, and file folders are named ambiguously.



Applications to SNSF Requirements

Data Management Plan (DMP)

Determining how files will be organized is also an important aspect of the DMP. You should indicate the structures of folders clearly in your DMP. An example of this might look like:

The project has a main, overarching folder that includes all related files and data. Within the main folder, each aim of the experiment has its own subfolder. Within an aim subfolder, data and related files are categorized by experiment type. An example hierarchy of folders is:

- *Dental Implant Project*
    - *Aim 1: Optimize Measurements*
        - *Depth*
            - *Procedure*
            - *Raw Data*
            - *Processed Data*
    - *Aim 2: Optimize Materials*
        - *Aluminum*
            - *Procedure*
            - *Raw Data*
            - *Processed Data*
        - *Steel*
            - *Procedure*
            - *Raw Data*
            - *Processed Data*

FAIR Data Principles

File structure organization is also important when ensuring your data and your project adheres to the FAIR Data Principle of being findable. Metadata and the data it describes should be easily searchable by both humans and computer systems in large indices of data. As stated previously, your files and folders should be structured simply and logically with a certain degree of specificity so that the data is easily differentiated. Well-organized file structures will allow you and others to quickly discover and identify interesting sets of data.

Conclusions

In summary, keeping a well-maintained file folder structure will allow you and other researchers to quickly and easily locate and identify files. With the proper organization, time spent looking for certain files and folders can instead be spent on research.

Chapter 5: Storing and Transferring Data Files

This section provides recommendations for storing and transferring data.

Any data that is generated from experiments or projects is highly valuable. Because of this, effectively storing and transferring the data is crucial.

Handwritten Data

The following are good rules of thumb for storing and transferring handwritten data:

- **Digitize any handwritten notes or hand-drawn pictures or diagrams.** Scanning or typing any handwritten notes and saving them onto a computer allows the information to be kept in two places in case the physical copy is lost. Typing handwritten notes will also improve comprehensibility, as handwriting can occasionally be illegible.
- **Have a safe, centralized, organized location for physical copies of notebooks or diagrams.** This allows the information to be found easily and efficiently. Additionally, this protects the physical copies from damage.
- **Keep physical copies in the lab.** Occasionally, experiments are conducted in sterile, controlled environments. Keeping the physical copies of notes or diagrams in the lab will decrease the risk of contamination. If the information is needed outside the lab, copy the page or diagram onto another sheet of paper or print out a digitized version.

Electronic Data

The following are good rules of thumb for storing and transferring electronic data:

- **Perform regularly scheduled back-ups.** This can be nightly, weekly, or monthly. Consistently backing up your data will prevent accidental loss.
- **Clear out unnecessary data.** Computers attached to instruments are not meant to be storage systems. Back-up any needed data onto the cloud or to your own computers and delete any files that are no longer needed. If you are a non-permanent lab member, delete any of your unnecessary project files, and make it clear to your colleagues as to which files are not being deleted and why.
- **Avoid transferring data via external drives, such as flash drives.** If external drives are broken or misplaced, whole sets of data from experiments or projects can be lost.
- **Communicate with your IT department.** Occasionally, the biggest obstacle to efficient data storage and transferal is a lack of communication between labs and IT. Talk to your IT department to determine the easiest way for you to transfer and store data.
- **Create duplicates if you plan to edit or manipulate a file.** Maintaining an original version of a file can be useful if you erroneously change or manipulate a file.

Applications to SNSF Requirements

Data Management Plan

Data Management Plans specifically require information on how the data will be stored, preserved, shared, and reused. In terms of completing the DMP, there are specific things you should think about and describe when conducting your project.

- **Access rights**
  At the time of the project, data should only be accessible by authorized persons. Ideally, any time a file or dataset is accessed, viewed, or changed during a project should be logged somewhere.
- **Storage and backup procedures**
  The frequency of data backups, physical or digital location of the storage medium, the network capabilities of the storage location, and related information should be mentioned.
- **Preservation procedures**
  The selection criteria for data preservation, length of preservation, and other relevant information as it pertains to preservation procedures should be detailed.
- **Sharing repositories and methods**
  The repositories on or methods through which data will be shared after publication should be well-documented, if applicable.

An example description for this information might look like the following:

Data access will be defined before starting the project, and the list of authorized peoples will be managed by the project head. Data is backed up nightly onto a restricted, centralized system provided by the university and managed by the IT department. Data will not leave the lab premises physically nor virtually unless authorized by the project head. Any data used for publication will be preserved for 10 years on the university server.

Research data is anonymized in accordance with institutional policy. Users are granted access with permission from the trial manager, access to any files will be logged and traceable. Data is backed up weekly onto a data center located in the facility, and after completion of the trial, will be stored for 20 years.

FAIR Data Principles

Your data's accessibility, interoperability, and reusability in accordance with the FAIR Data Principles will depend heavily on how you choose to transfer and store your data.

For accessibility and reusability, the SNSF at the very least expects you to share the metadata needed to make any published results reproducible. If you choose to share data, it can be in raw or processed form but should be accompanied with relevant metadata. If there are any particular tools or software needed to reuse, reproduce, or access the data, this information should be documented, and the tools or software used should be made available. Ideally, at least the metadata should be accessible by anyone with a computer. If necessary, the data can be restricted by authentication or authorization procedures.

For interoperability, your data and metadata should be machine-readable, or in a form that a computer can process. Additionally, you should properly link and cite any relevant files or datasets and describe their relevance.

If these requirements cannot be met, you should explain the reasoning in detail in your DMP.

Conclusions

Properly storing and transferring research data is important to its eventual analysis and publication. By adhering to the practices stated, you can be assured your data is not lost.

Additional References

Further information concerning SNSF guidelines and data management can be found through the following websites:

The DCLM Project – The Swiss “Data Life-Cycle Management” (DCLM) project aims to provide researchers with guidelines on research data management. An example DMP template and additional information concerning these guidelines can be found at https://www.dlcm.ch/

Swiss National Science Foundation (SNSF) – The Swiss National Science Foundation has provided a document explaining the contents of the mySNF form and the Data Management Plan, which can be found here: http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf

Additionally, a thorough explanation of the FAIR Data Principles can be found at http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Lastly, the language policy of the Swiss National Science Foundation can be found here: http://www.snf.ch/SiteCollectionDocuments/Dossiers/dos_sprachenpolitik_snf_e.pdf

National Institutes of Health (NIH) – The National Institutes of Health has provided guidelines for scientific record keeping, which can be found here: https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-scientific_recordkeeping.pdf

The University of Cambridge – The University of Cambridge has provided useful information on general data management, which can be found here: https://www.data.cam.ac.uk/data-management-guide

Stem Cell Technologies – Stem Cell Technologies has provided useful information on how to organizing and document lab notebooks, which can be found here: https://www.stemcell.com/efficient-research/lab-notebooks-references-protocols

Paper Lab Notebook Checklist

Have you...	✓, X, or N/A
... started on a new page?	
... made an associated table of contents entry?	
... dated the page?	
... recorded the name of the experiment or project?	
... recorded the aim or purpose of the experiment or project?	
... recorded the procedure in detail?	
... recorded experimental conditions in detail?	
... recorded observations and results in detail?	
... explained and expanded any abbreviations?	
... explained any mistakes or failed results?	
... attached any external graphs or computer printouts?	

Electronic Note-Taking Checklist

Have you...	**✓, X, or N/A**
... created a note-taking template if this a common experiment or project?	
... shared the template with colleagues for widespread use, if applicable?	
... typed in the date?	
... recorded the name of the experiment or project?	
... recorded the aim or purpose of the experiment or project?	
... recorded the procedure in detail?	
... recorded experimental conditions in detail?	
... recorded observations and results in detail?	
... explained and expanded any abbreviations?	
... explained any mistakes or failed results?	
... imported any relevant images and made annotations?	
... named the document file with descriptive keywords?	

Audio Note-Taking Checklist

Have you...	✓, X, or N/A
... tested the recording device to ensure you can be heard?	
... stated the date?	
... stated the name of the experiment or project?	
... stated the aim or purpose of the experiment or project?	
... stated the procedure in detail?	
... stated experimental conditions in detail?	
... stated observations and results in detail?	
... explained any abbreviations?	
... explained any mistakes or failed results?	
... verbally related the audio recording to any relevant notes (handwritten or electronic) that are taken concurrently?	

Naming Files Checklist

Have you...	✓, X, or N/A
... included the date?	
... included your name and the name of any collaborators?	
... included the experiment type?	
... included the trial number?	
... included specific, descriptive keywords?	
... communicated the standards for labeling files with the project team?	
... created or updated a glossary of abbreviations and their expanded forms?	

Organizing Data Files Checklist

Have you...	✓, X, or N/A
... created a main folder for the project?	
... created subfolders for the different stages, tasks, and aims of the experiment?	
... identified what additional subdirectories you will need?	
... grouped folders by date?	
... discussed folder structure standards with any collaborators?	
... created shortcuts to frequently accessed folders?	
… moved unnecessary folders out of the desktop?	

Storing and Transferring Physical Data Checklist

Have you...	✓, X, or N/A
... digitized the data?	
... identified where the data will be stored in a central location?	
... communicated the storage location to any collaborators?	
… communicated whether or not the data can be taken out of the lab space?	