



UNIVERSITÉ
DE GENÈVE

BIBLIOTHÈQUE



PhD Library Camp

Module 1




Managing research data as a junior scientist

TABLE OF CONTENTS

Table of contents	2
Learning outcomes	2
Open science principles.....	3
Data management introduction.....	3
Data organization	4
Data storage	6
Electronic Laboratory Notebooks (ELNs).....	7
Data repositories	7
Creative commons licenses.....	9
Data management plan (DMP)	10
Documentation	11

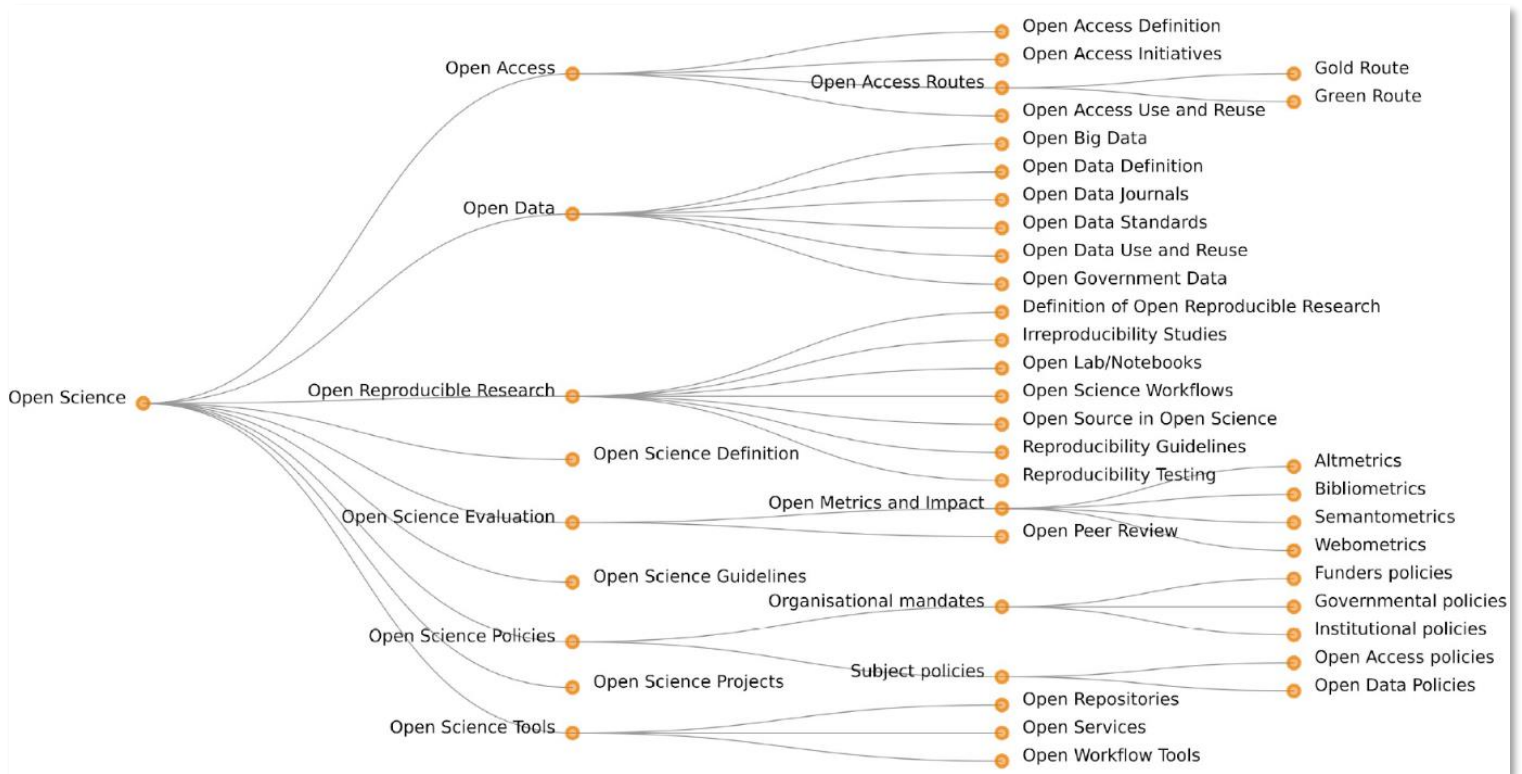
LEARNING OUTCOMES

By the end of this workshop, you will be able to:

-  Apply good practices to manage research data
-  Value the advantages of depositing your research data in a repository and how to select it
-  Anticipate the management of your research data and complete a Data Management Plan (DMP)

PART I

OPEN SCIENCE PRINCIPLES



Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). Fostering Open Science to Research Using a Taxonomy and an eLearning Portal. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business* (p. 11:1–11:8). New York, NY, USA: ACM. <https://doi.org/10.1145/2809563.2809571>

DATA MANAGEMENT INTRODUCTION



Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). Fostering Open Science to Research Using a Taxonomy and an eLearning Portal. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business* (p. 11:1–11:8). New York, NY, USA: ACM. <https://doi.org/10.1145/2809563.2809571>

Managing properly your data at all steps of your project is crucial and you will benefit from it for many reasons:

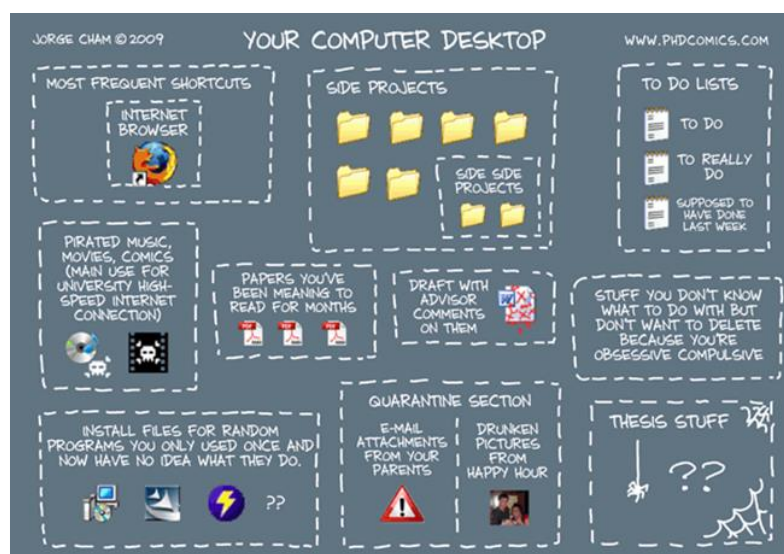
- **Increase your research impact:** by sharing your data, your number of citations and future collaborations will potentially increase
- **Save your time:** planning ahead for your data management needs will save you time and resources.
- **Preserve your data:** depositing your data in a repository safeguards your investment of time and resources while preserving your research contribution for you and others to use.
- **Maintain data integrity:** managing and documenting your data throughout your project maximises transparency and accountability.
- **Meet grant requirements:** many funding agencies now require that researchers deposit data collected as part of a research project.
- **Facilitate new discoveries:** sharing your data promotes innovation and potential new data
- **Support open access:** spread the knowledge and allow the knowledge to be built on. Show your support for open access by sharing your data.

DATA ORGANIZATION

Organizing your files will help you and your colleagues to find easily now and in the future your data.

For your folders organization, use the hierarchical method, using the following tips:

- A. Avoid overlapping categories**
- B. Don't let your folders get too big**
- C. Don't let your structures go too deep**
- D. Order dates beginning with the year to enable sorting by date (e.g., YYYYMMDD)**
- E. Embed metadata in your files (if possible)**
- F. Add shortcuts to files within other relevant folders**



Jorge Cham, 2009. <http://www.phdcomics.com/comics/archive.php?comicid=1168> [Accessed January 9 2017]

Name your files according to conventions, such as:

Best Practice	Example
Order dates beginning with the year to enable sorting by date	YYYYMMDD
Limit the file name to 32 characters	32CharactersLooksExactlyLikeThis.csv
For sequential numbering, use leading zeros to allow for multi-digit versions For a sequence of 1-10: 01-10 For a sequence of 1-100: 001-010-100	NO : ProjID_1.csv ProjID_12.csv YES : ProjID_01.csv ProjID_12.csv
Don't use special characters & , * % # ; * () ! @ \$ ^ ~ ' { } [] ? < > -	NO : name&date@location.doc
Use only one period (for the file extension)	NO : name.date.doc YES : name_date.doc
Avoid using generic data file names that may be ambiguous when moved	NO : MyData.csv YES : ProjID_date.csv
Keep track of document versions sequentially (e.g. v01, v02)	NO : Thesis_final.doc YES : Thesis_v03.doc



http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf

Some **tools** may be useful to rename your files:

- Tools on your computer:
 - Finder or Automator (Mac)
 - Explorer and Powershell (Windows)
- Dedicated software
 - Ant Renamer (Windows, open source, free)
 - PSRenamer (Windows, Mac or Linux, open source, free)

DATA STORAGE

- You should store three copies:
 1. the original
 2. a copy kept on a local external device
 3. a copy kept on an external device at a different location
- Include metadata and readme.txt files in your backups.

10 simple rules for data storage

(from : Hart EM, Barnby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten Simple Rules for Digital Data Storage. PLOS Computational Biology. 2016;12: e1005097. doi:10.1371/journal.pcbi.1005097)

1. Anticipate How Your Data Will Be Used
 2. Know Your Use Case
 3. Keep Raw Data Raw
 4. Store Data in Open Formats
5. Data Should Be Structured for Analysis
6. Data Should Be Uniquely Identifiable
 7. Link Relevant Metadata
8. Adopt the Proper Privacy Protocols
9. Have a Systematic Backup Scheme
10. The Location and Method of Data Storage Depend on How Much Data You Have

Quality data storage platforms :

A. NAS of UNIGE: the Network-attached Storage is the system dedicated to UNIGE research data storage. NAS files are stored daily and duplicated every 4 hours to a UNIGE back-up NAS located on the Biotech campus. Storage costs CHF 100/TB/year: <https://www.unige.ch/researchdata/en/store/store-menu/storage/>

B. SWITCH Drive: swiss universities platform. Free until 50 GB of data. Possibility of sharing your folders: <https://www.switch.ch/drive/>



ELECTRONIC LABORATORY NOTEBOOKS (ELNs)









ELNs are more and more used by scientists as an alternative to traditional paper-based notebooks.

Advantages of ELNs are:

- 👍 Text-based searching
- 👍 Integration of **metadata** (data *about* the experiment like date, temperature...) → facilitates retrieval, analysis and reuse
- 👍 Insertion of all types of data and media (images, movies, protein sequences...)
- 👍 Customized templates to facilitate data entry
- 👍 Facilitated data management among an entire lab
- 👍 Archives can be kept in the lab after people are leaving

But using an ELN could have a cost (fees for the software and time to learn to use it) and requests to define clear access rights.

Examples of ELN

LICENSED ELN	OPEN SOURCE ELN
	
	
	
	

DATA REPOSITORIES

Open data repositories are digital infrastructures that allow researchers to store, share, access and preserve for the long term their research data. Some are generic, other disciplinary or institutional. They can help to improve the dissemination and reuse of scientific data. If you wish your data to be cited or reused by others or if it is requested by your publisher, you may share your scientific data sets in such repositories.

FAIR principles

In order to share your data and allow other people to reuse them, you should try to have your data compliant with the FAIR principles:

- ✓ **F**indable: (meta)data are assigned a globally unique and persistent identifier
- ✓ **A**ccessible: (meta)data are retrievable by their identifier using a standardized communications protocol
- ✓ **I**nteroperable: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- ✓ **R**eusable: meta(data) are released with a clear and accessible data usage license

There are several repositories:

- A. Yareta** is the institutional repository from UNIGE, created in 2019. All UNIGE members can use it for free up to 50 GB.



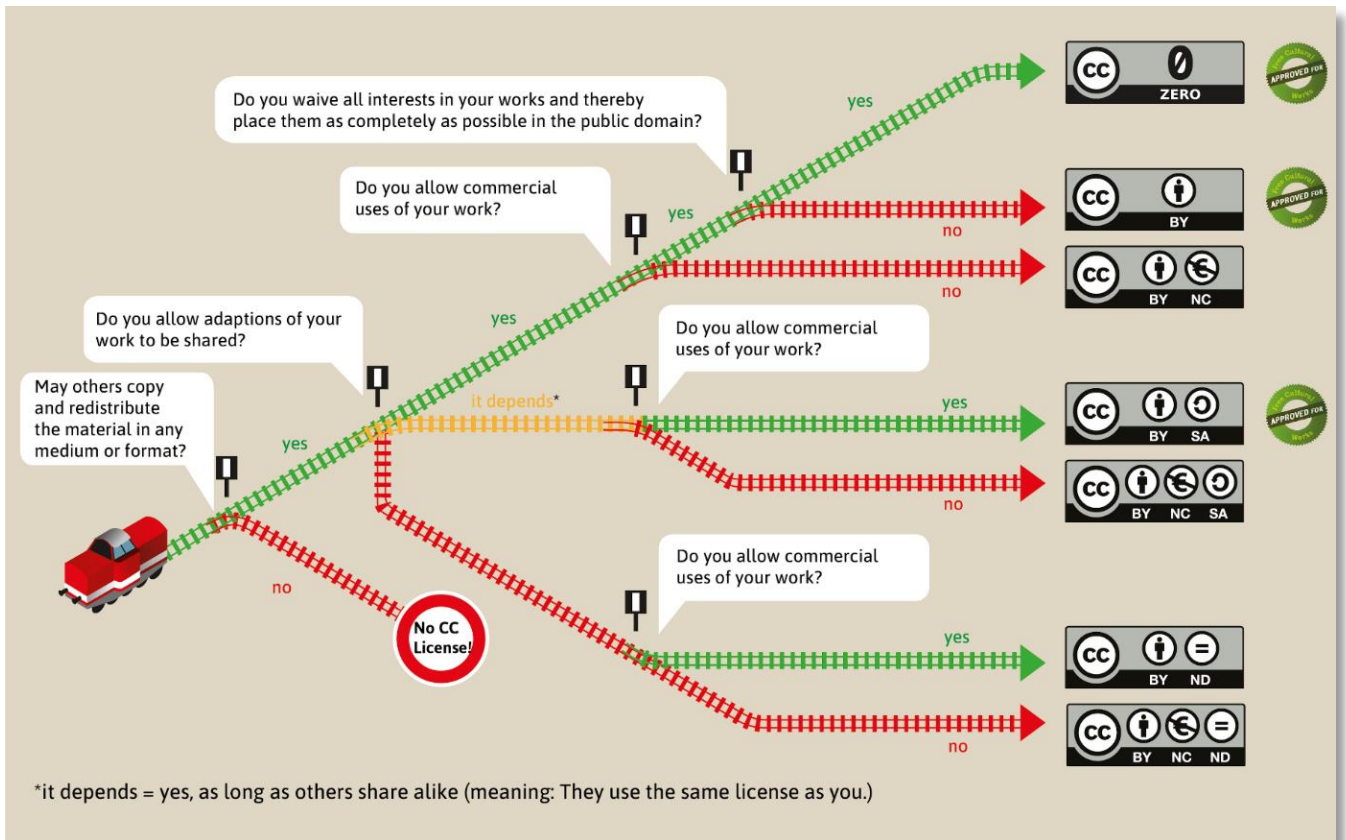
- B. Zenodo** is a repository created by OpenAIRE in 2013 and based in CERN. It's a « general-purpose » repository that allows files up to 50GB. There are no uploading fees.



- C. Re3data.org (Registry of Research Data Repositories)** is a registry of all existing repositories. You can search for domain-specific or general purposes registries.



CREATIVE COMMONS LICENSES



This Graphic "Choo-Choo-Choose your license!" is based on the work "Welche CC-Lizenz ist die richtige für mich?" by Barbara Klute und Jöran Muuß-Merholz für wb-web* unter CC BY-SA 3.0**. The English version is a translation and enhancement by Jöran Muuß-Merholz under the same licence.
 * <http://www.wb-web.de> | ** <https://creativecommons.org/licenses/by-sa/3.0/deed.de>

PART II: WRITE YOUR DMP

DATA MANAGEMENT PLAN (DMP)

A Data Management Plan (DMP) is a formal document that outlines how the data of a research are to be handled both during a research project, and after the project has ended. It is a living document that needs to be updated throughout the research project.

The Swiss National Foundation (SNF), and many other funding agencies, require a DMP to be filled when applying for a grant.



Here is the form from the SNF:

1 Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2 Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3 Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4 Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 I will choose digital repositories that are conform to the FAIR Data Principles. [CHECK BOX]
- 4.4 I will choose digital repositories maintained by a non-profit organisation. [RADIO BUTTON yes/no]

Detailed content of the SNF Data Management Plan:

http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf

General purpose

Graduate information literacy module. Available from: <http://www.informationliteracy.ie> [Accessed January 9 2017]

Data organization

CNET. *Two Ways to Batch Rename Files on a Mac*. Available from: <https://www.cnet.com/how-to/two-ways-to-batch-rename-files-on-a-mac/> [Accessed August 15 2018]

MAKE TECH EASIER. *3 Ways to Batch Rename Files in Windows*. Available from: <https://www.maketecheasier.com/batch-rename-files-in-windows/> [Accessed August 15 2018]

ANT RENAMER. Available from: <https://antp.be/software/renamer/fr> [Accessed August 15 2018]

PSRENAMER. *PSRenamer | PowerSurge Publishing*. Available from: <https://www.powersurgepub.com/products/psrenamer/index.html> [Accessed August 15 2018]

Data Management Plan (DMP)

Research Data - Research Data - UNIGE [Internet]. 19 Oct 2016 [cited 9 Aug 2019]. Available: <https://www.unige.ch/researchdata/en/>

Data Management Plan (DMP) - Guidelines for researchers - SNF [Internet]. [cited 9 Aug 2019]. Available: http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx

Dlcm template for the snsf data management plan :: dlcm [Internet]. [cited 27 Jan 2020]. Available: <https://www.dlcm.ch/blog/dlcm-template-snsf-data-management-plan>



Bibliothèque de l'UNIGE, 2020

This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License: <http://creativecommons.org/licenses/by-sa/4.0/deed>.