FORS data management webinar series

Data Management Planning

Dr. Alexandra Stam

September 9th, 2021

FORS webinar series

1. Introduction to data management planning

2. **Research ethics** / September 30th

3. **Data security** / October 12th

4. **Data sharing** / November 4th

FORS

Presentation outline:

1. About FORS
2. Open research data
3. Research data management
4. Data management planning
5. Depositing data with FORS

About FORS

FORS – Swiss Centre of Expertise in the Social Sciences

- Research infrastructure of national scope intended for any institution or person active in the social sciences
- Funded by the Confederation and the Swiss National Science Foundation (SNSF), hosted by the University of Lausanne
- Founded in 2008 by merging existing entities (SIDOS, SHP, Selects)

What we do

Infrastructure	Project catalogue, data archiving and dissemination service, support
Data production	National (SHP, Selects) and international surveys (ESS, MOSAiCH-ISSP, SHARE, EVS)
Research	Thematic, methodological, and epistemological research
Collaboration	At the national and international level, in all our active domains
Development	Cutting-edge tools, data collection, and data linking

Archiving service

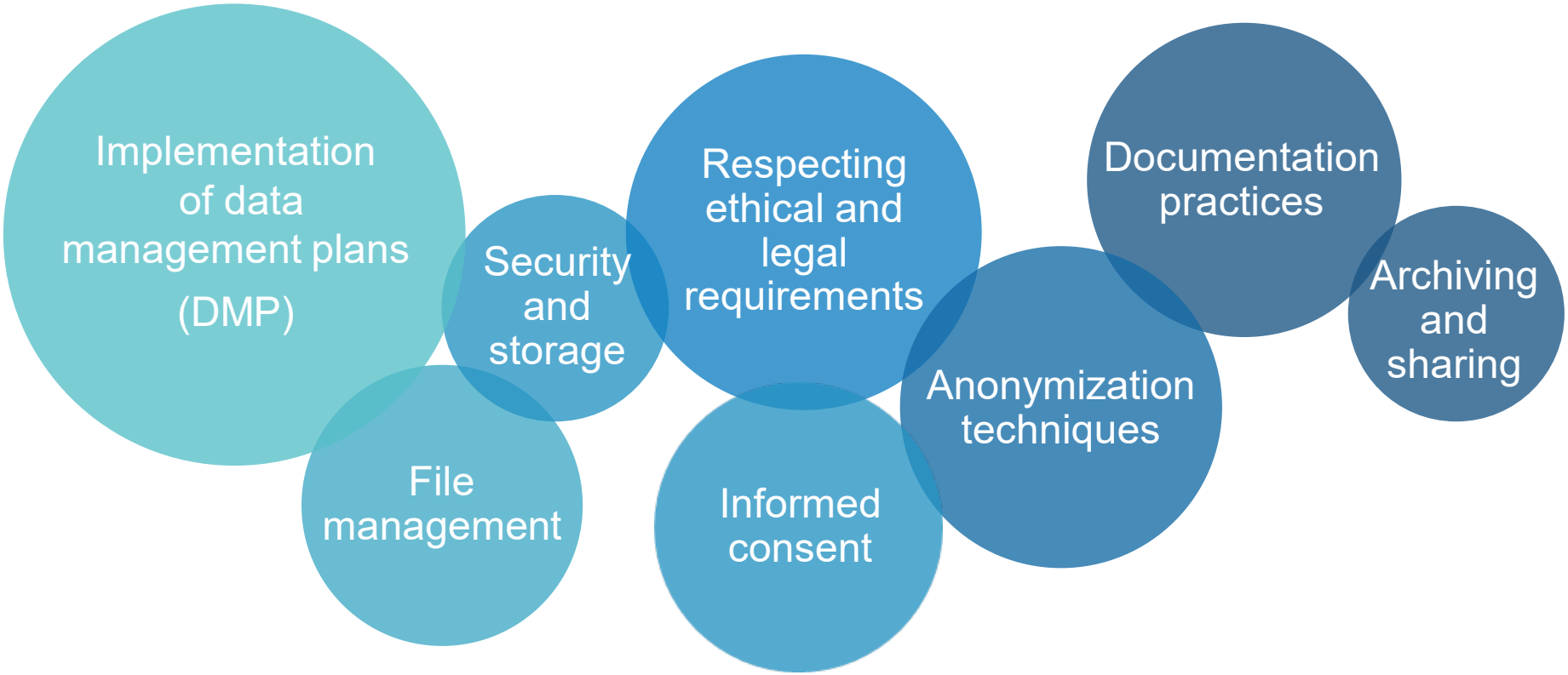
FORSbase

Data archiving	Data access	Data management
New requirements	Direct access to:	Training
Long-term preservation	+ 700 datasets	Consultancy
Enhance the value of research projects	+ 11’000 project descriptions	Development of materials (e.g., guides)

Our digital repository FORSbase will soon become

SWISSbase

Data management support



FORS

Open research data

Definition of open data

"*Open research data refers to the data underpinning scientific research results that has no restrictions on its access, enabling anyone to access it.*" [European Commission]

It follows three rationales:

- Ideological: 'data as a public good'
- Scientific transparency: 'data replication'
- Economic: 'data re-use'

New requirements

from funders:

- Data management plans (DMPs)
- Data sharing (in FAIR repositories)

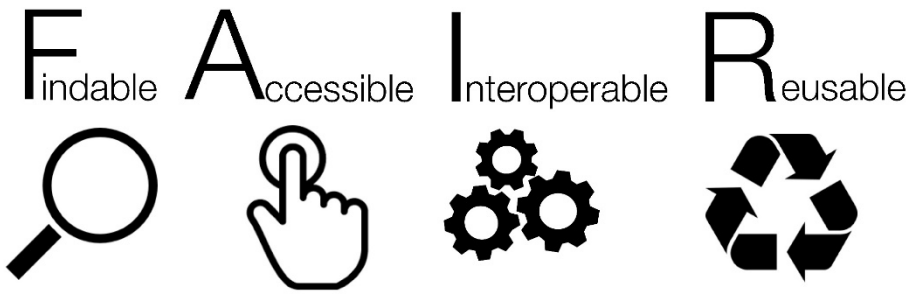
from institutions:

- Data management plans (DMPs)
- Data sharing

from journals:

- Deposit of data used in publications
- Sufficient documentation

FAIR principles



In practice: ‘accessibility’ rather than ‘openness’

“*This is a key, but often misunderstood element of FAIR. The ‘A’ in Fair does not necessarily mean ‘open’ or ‘free’. Rather, it implies that one should provide the exact conditions under which the data may be accessed. Hence, even heavily protected and private data can be FAIR*” (www.go-fair.org/fair-principles)

Some advantages to data sharing

On top of facilitating the reproduction and verification of research results, as well as allowing data re-use, data sharing:

- makes research work and results more visible;
- increases the number of citations of scientific articles for which research data is also published;
- encourages new collaborations and new avenues of research;
- meets the requirements of some scientific funders and journals.

Main challenges for the social sciences

- Subjective: «qualitative data cannot be shared»
- Practical: lack of know-how (data protection techniques)
- Normative: tension between data openness and data protection laws

Challenges are especially important when it comes to personal and sensitive data

Research data management

Definition of research data management

Research data management includes all activities associated with data other than the direct collection and use of the data.

It covers all aspects of handling, organising, documenting and enhancing research data, and enabling their sustainability and sharing.

Some key data management skills for research:

- data and project planning;
- data collection considerations (e.g. informed consent);
- data preparation;
- documentation;
- anonymisation;
- data organisation;
- data storage and security;
- dissemination;
- copyright; and
- data sharing.

Why is research data management important?

Advantages of good data management go beyond meeting open data requirements. It also saves significant time and improves research and data quality. For example it:

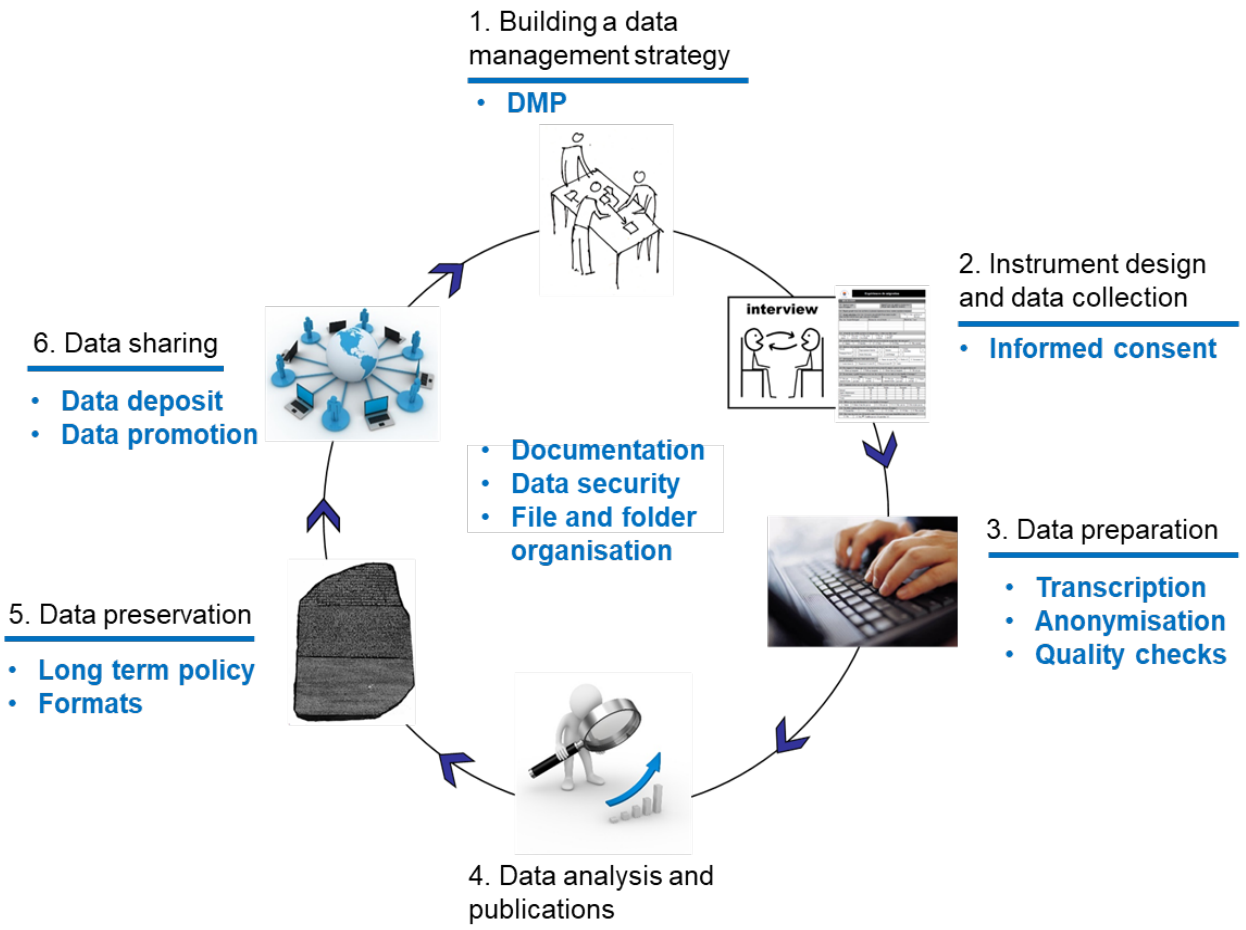
- allows you to be more transparent about how you have generated and transformed your data;
- allows your colleagues to re-use your data;
- allows you to come back to your data further down the line;
- allows you to make fewer mistakes.

Research data lifecycle

Data management needs to be considered throughout your project:

1. At the start of your project, while planning how you intend to create, process and preserve your data;
2. During the project, where you manage your data on a day-to-day basis;
3. At the end of the project, to preserve the data and make them available whenever possible.

Example of the data lifecycle



Planning versus day-to-day data management

Data management planning	vs	Day-to-day data management
▪ general overview ▪ generally rather brief ▪ intentions of good practices ▪ strong focus on data sharing ▪ expected ‘problems’		▪ applied data management ▪ detailed strategy ▪ clear rules ▪ focus on immediate needs throughout the life-cycle ▪ actual solutions
e.g., DMP		e.g., fixing rules; drafting a consent form

Data management planning

Definition of a DMP

A DMP is a document that lays out how you intend to manage your research data. It will normally explain how you will organise the creation, processing and preservation of your data during the life of your project and beyond.

Some advantages

- Provides an overview of the steps to be taken in your project
- Strong incentive to make data management decisions before they could potentially become a problem (security, consent, etc.)
- Can be used to make strategic decisions (ex. data sharing strategy)
- It's a great tool to define procedures when working in a team
- Allows for meeting funder or institutional requirements

It’s more than an administrative burden!

SNSF requirements and review process

- DMPs are mandatory for each research proposal;
- DMPs concern all data collected and produced in the course of the research project;
- A proposal cannot be submitted as long as the DMP has not been completed online;
- DMPs are not part of the review process;
- The DMP is meant to be a living document;
  - It can/should be adapted throughout the project
  - A final version needs to be submitted at the end of the project
- At least data underlying a publication must be shared;
- The SNSF recognises exceptions to sharing.

Content of the SNSF DMP

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?

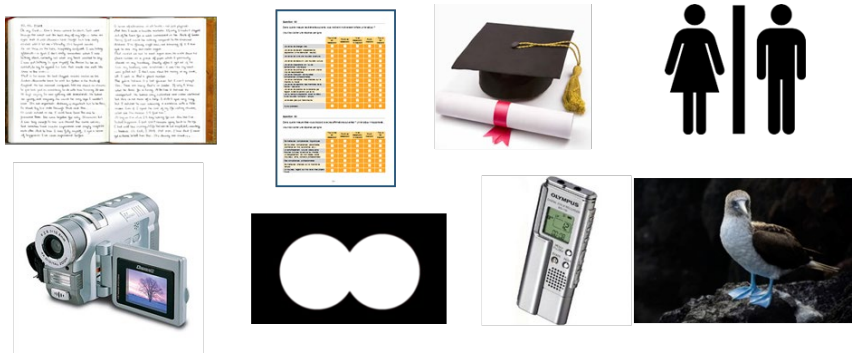
4. Data sharing and reuse

- How and where will the data be shared?
- Are there any necessary limitations to protect sensitive data?
- [checkbox: I will choose digital repositories conform to the FAIR data principles]
- [Yes/No button: I will choose digital repositories maintained by a non-profit organisation]

1. Data collection and documentation

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

Main purpose: provide a description of the data that will be produced/collected/used/reused in your research project. Data may be defined according to their type, degree of intervention of the researcher in their production, level of processing, format, or legal status.



Special attention must be given to personal and sensitive data!

How will the data be collected, observed or generated?

Main purpose: provide a description of the methods and instruments of data collection used. It is also a question of clarifying the strategies for organizing these data.

Instruments / methods / quality control: the possibility or not of avoiding unnecessary collection of personal or sensitive data depends largely on the types of methods used.

Organisation: It is a matter of demonstrating that a data organization strategy has been considered.

What documentation and metadata will you provide with the data?

Documentation is any information that serves as a record of a research project and that renders data usable and meaningful. Documentation in the social sciences describes the context of the study, the research process, how the data were collected and manipulated, and the structure and content of the data.

Metadata are data about the data. In the social sciences it consists in descriptors that facilitate cataloguing data and data discovery. Unlike documentation, metadata is a formally agreed upon set of standards often with controlled fields and vocabularies.

2. Ethics, legal and security issues

How will ethical issues be addressed and handled?

Main purpose: provide information on how participants' rights and integrity will be protected during and after the project. It must indicate:

- What are the risks (physical and/or psychological) faced by participants when taking part in the study;
- What standards apply to the types of data generated by your project;
- What measures are planned to obtain the necessary authorisations for the collection, processing and sharing of personal/sensitive data;
- What are the measures planned to protect personal and sensitive data.

How will data access and security be managed?

Main purpose: describe the measures planned to prevent any illegitimate access to the data during production, processing and storage, as well as those provided to prevent any damage or loss. Some security risks include:

- hardware failure; software malfunction; degradation of storage media
- accidental or malicious damage/modification to data
- theft of data
- natural disaster or fire
- inappropriate access

How will you handle copyright and intellectual property rights issues?

Main purpose: clarify data ownership and rights

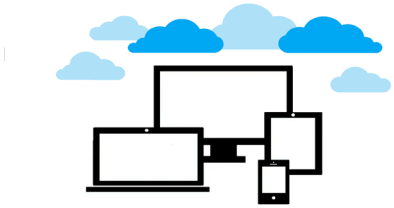
3. Data storage and preservation

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

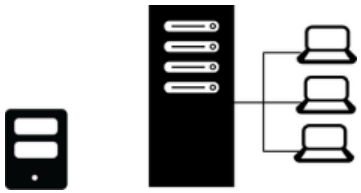
Main purpose: describe where your data will physically be stored during the research process and how you will protect them from loss.

Local devices

Cloud storage



Network drives



What is your data preservation plan?

Main purpose: 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. Ask yourself the following questions:

- What data to I want to keep beyond the research project?
- How long do I want to keep the data and why?
- How will I ensure data preservation? Where will I store my data and how will I make sure my data remain accessible over time? Will any of my data be deposited in a data repository or data archive?

4. Data sharing and reuse

How and where will the data be shared?

FAIR and non-commercial repository

Different solutions, some better than others

- (trusted) domain-specific repository
- institutional repository;
- general purpose repository (e.g., Zenodo, Figshare);
- journal supplementary material services;

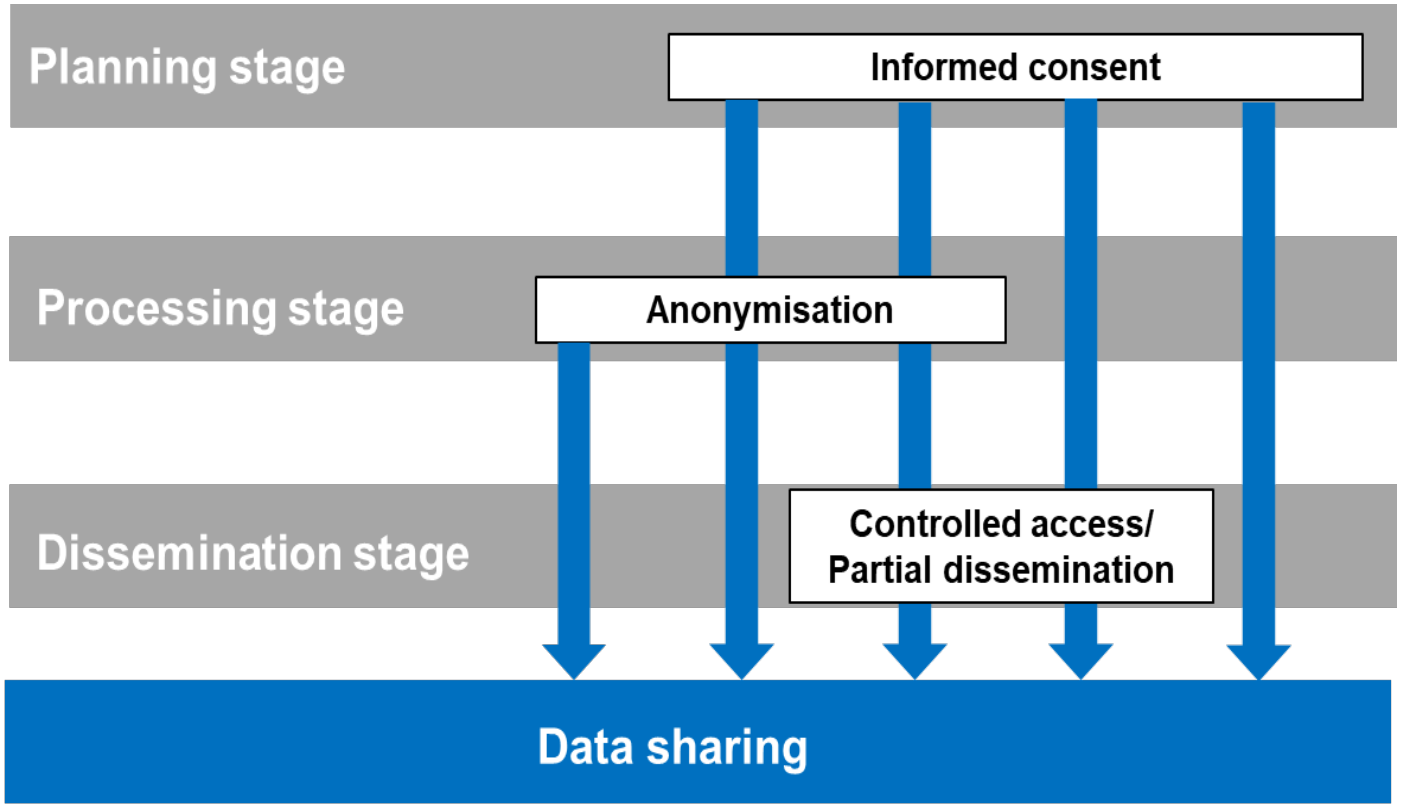
A full list of repositories can be found on Re3data.org

Are there any necessary limitations to protect sensitive data?

Main exceptions to data sharing

Legal	Ethical	Copyright	Confidentiality
Example: sensitive, non-anonymised data for which no consent has been obtained.	Example: Data for which consent has been obtained but where the risks to participants is too high.	Example: Data that have been produced by someone else and are legally protected	Example: Data that have been obtained on the condition that they remain confidential.

Special conditions for sensitive data



Depositing data with FORS

Some advantages of sharing with FORS

Go with a trusted Swiss solution

- ✔ 100% Swiss, in line with Swiss legal requirements, data on Swiss servers, Swiss quality
- ✔ FAIR repository recognized by the Swiss National Science Foundation
- ✔ Curated by archive experts
- ✔ International certification of quality (CoreTrustSeal)

Increase your visibility

- ✔ Get free persistent identifiers (DOIs) for all published datasets to increase your citation rates
- ✔ Get assistance with documenting your data according to international, field-specific standards
- ✔ Benefit from the visibility, outreach, and international exposure of FORSbase, the main data repository for the social sciences in Switzerland

Protect your rights and keep control over your data

- ✔ Enjoy long-term preservation of your data – not limited in time
- ✔ Access control options (your prior approval, embargo, restrictions on use)
- ✔ Protection of depositors' rights (binding deposit and user contracts)
- ✔ Access statistics on data downloads and the use of your data

Take advantage of our personalized services throughout the research lifecycle

- ✔ We help you find and access the data and research information you are looking for
- ✔ We assist with the deposit and sharing of your data
- ✔ We provide data management consulting
- ✔ Our services are free of charge

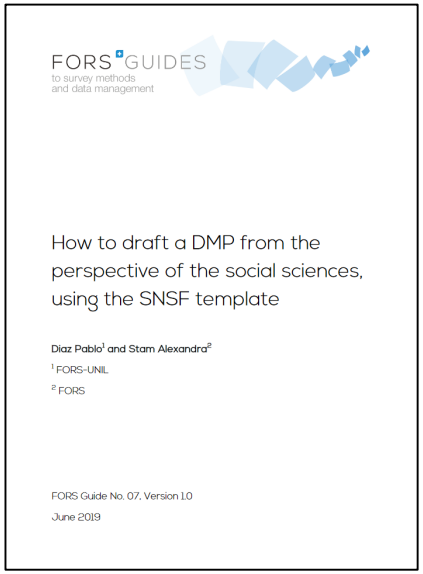
FORS

Resources

Data management guides

- Ethics in the era of open research data;
- Informed consent;
- How to draft a DMP
- Pre-registration and registered reports
- Data anonymisation, legal, ethical and strategic considerations

https://forscenter.ch/publications/fors-guides/



Archiving resources

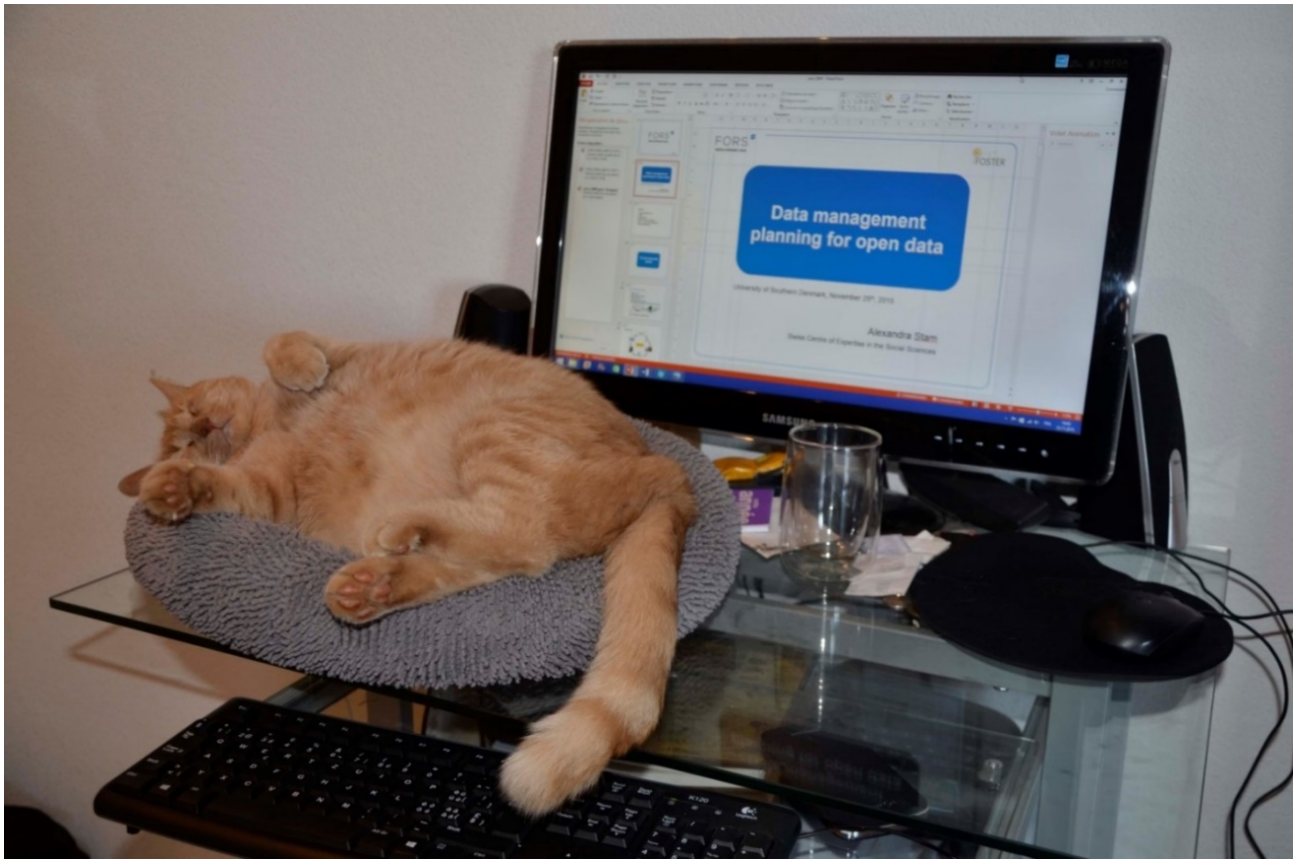
https://forscenter.ch/data-services/help-resources/

International resources

CESSDA data management expert guide

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide

Questions?



Contact: rdmservices@fors.unil.ch