

Introduction to Data Management

Doctoral school - linguistics
Macolin, 10.12.2018

Alexandra Stam
Pablo Diaz

Outline

1. Presentation of FORS

- FORS main activities
- Data management at FORS

2. Definitions

- Research data
- Data management

3. Why manage your data?

- Open data movement
- Meet funder and journal requirements
- Legal/ethical requirements
- For good science

4. How to do data management?

- What data do I have / will I be collecting?
- From planning to practice

1. Presentation of FORS

FORS main activities

Swiss Centre of Expertise in the Social Sciences

- Methodological research
- Large Surveys
- Data and research information services (DARIS)

DARIS

FORSbase

Data archiving

New requirements

Long-term preservation

Enhance the value of research projects

Data access

Direct access to:

+ 500 data sets

+ 11'000 project descriptions

Data management

Training

Consultancy

Development of materials (i.e. guides)


The screenshot shows the FORSbase website homepage. At the top left is the FORSbase logo with a database icon. To the right are language options: FR, EN, DE. Below these are buttons for 'Catalogue' and 'Home'. Further right are 'Sign up' and 'Login' links. The main content area features a large 'WELCOME TO FORSBASE' heading with a database icon, followed by the tagline 'Your research information and data access portal in Switzerland.' Below this is a paragraph explaining the platform's purpose and a link to a 'FORsbase: overview' guide. On the right side, there is a 'Useful Resources' section with a list of links: Quantitative data preparation for depositing, Qualitative data preparation for depositing, List of accepted file formats, Data documentation glossary, Deposit contract, User contract, FAQs, and Contact. Below this is the FORS⁺ logo and a link to 'Browse data on Nesstar'. At the bottom, there are four main action cards: 'Explore the catalogue' (with a binoculars icon), 'Get data' (with a shopping cart icon), 'Publish your study' (with a download icon), and 'Deposit your data' (with a person and plus icon). Each card contains a brief description and a link to a guide.

FORSbase FR EN DE

Catalogue Home

Sign up
Login

WELCOME TO FORSBASE



Your research information and data access portal in Switzerland.

FORSbase is an online platform that enables you to access data and obtain information about social science studies in Switzerland. You can register your own research projects, as well as store and share your own data.


Select one of the options below or first discover FORSbase's different functionalities in our guide [FORsbase: overview](#).

Useful Resources

- Quantitative data preparation for depositing
- Qualitative data preparation for depositing
- List of accepted file formats
- Data documentation glossary
- Deposit contract
- User contract
- FAQs
- Contact


FORS⁺

[Browse data on Nesstar](#)

 Explore the catalogue


Our collection includes over 10'000 social science research projects and is constantly updated. Discover the catalogue and liaise with the researchers.

[Guide: Explore the catalogue](#)

 Get data


Discover and instantly download for free hundreds of datasets for your own research projects. Start your analyses right away.

[Guide: Get data](#)

 Publish your study

Register your research project in our database to gain visibility, expand your network, and develop new collaborations.

[Guide: Publish your study](#)

 Deposit your data

Share your data and ensure their long-term preservation while meeting funder and/or journal requirements and increasing citations of your research.

[Guide: Archive your data](#)

Data management at FORS

Brian Kleiner



PhD in linguistics
Qualitative and quantitative
methods

Pablo Dias



PhD in political sociology
Qualitative methods

Alexandra Stam



PhD in geography
Qualitative and quantitative
methods

Marieke Heers



PhD in educational sociology
Quantitative methods



Our data management vision and strategy

Early days: Focus on DM from a data service point of view

- «We want your data» perspective
- «We want you to have fun, and apply good practices» perspective



And then came the Swiss Federal Survey of Adolescents



- We got to run a large-scale survey
- We wanted to make it the best survey ever (with respect to DM)
- We failed to apply the recommendations we were teaching (lack of tools and in-depth practical guidance)

Our data management vision and strategy


Early days: Focus on DM from a data service point of view

- «We want your data» perspective
- «We want you to have fun, and apply good practices» perspective



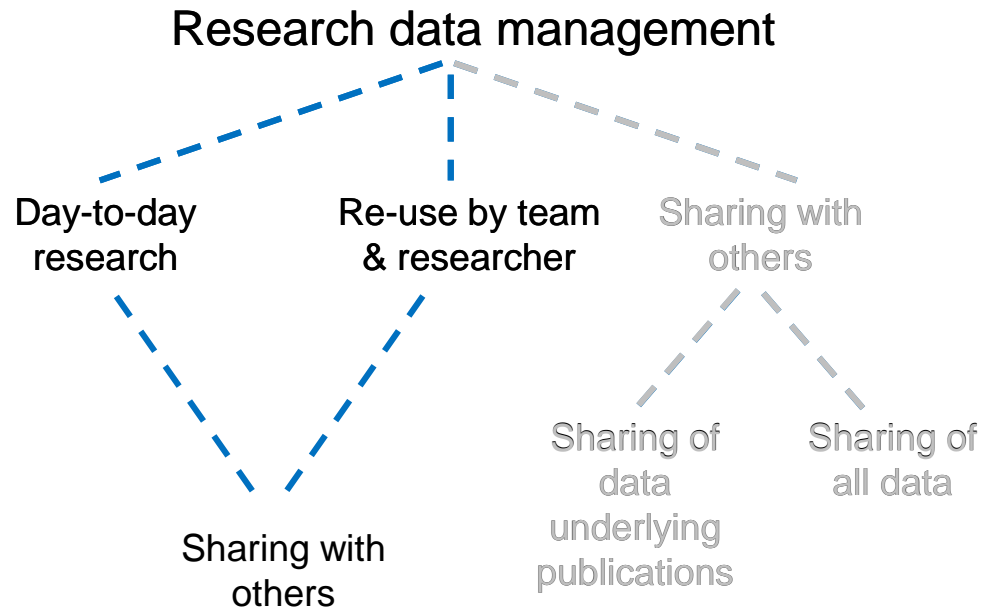
Current days: Focus on DM from a researcher's point of view

- «We want you to be able to apply good DM» perspective
- «We want you to see the value of DM for your work above all» perspective



It does not exclude data sharing or fun

Our vision



2. Definitions

Research data



written notes

survey data



visually recorded data



observational data



Human subject data



audio recorded data



pictures



Corpus data

*‘The material underpinning a research assertion’
(University of Sheffield, 2013)*

‘Research data can be extremely diverse: from spreadsheets, audio-visual materials, databases, to 3D-models and result lists from large experiments. Sizes may vary from a couple of small files related to a specific publication [...] to vast collections of experimental results (‘big data’), that can only be processed using specialized programmes.’ Source: <https://www.openaire.eu/>

Sensitive and/or personal data

Personal data are any information that enables direct or indirect identification of a human subject.

Although definitions may change across cultures and legal bases, personal data are usually considered sensitive when they relate to the following topics:

- Racial or ethnic origin;
- Political opinions;
- Religious or philosophical beliefs;
- Trade union membership;
- Physical or mental health
- Sex life
- Criminal offences and court proceedings;
- Genetic data;
- Biometric data.

Data management

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.

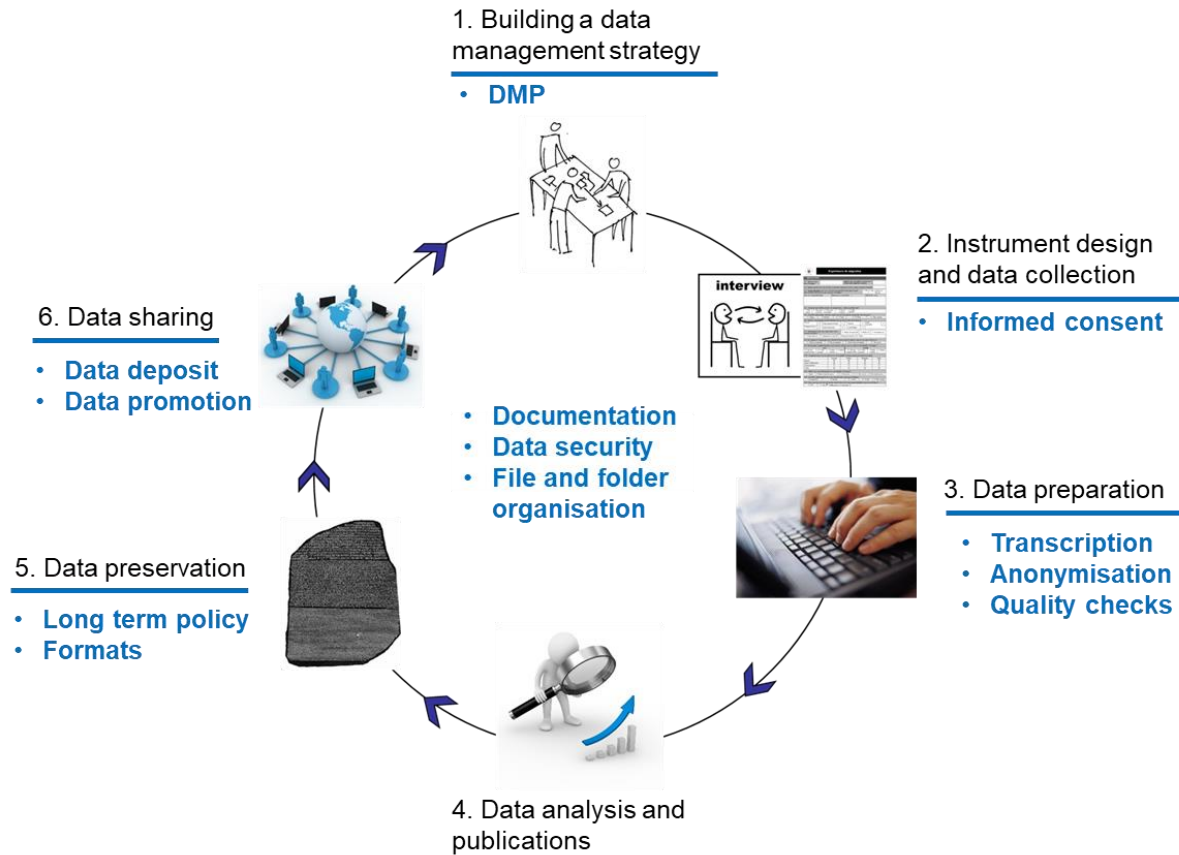
Good data management practices are important for:

- Day to day project data management
- Future uses of the data

Some key data management skills for social science research:

- Data and project planning
- Data collection considerations (e.g. informed consent)
- Data preparation
- Documentation
- Anonymisation
- Data organisation
- Data storage and security
- Dissemination and copyright
- Data sharing

Research life-cycle



Data management planning

- General overview
- Generally rather brief
- Intentions of good practices
- Strong focus on data sharing
- Expected 'problems'

i.e. DMP

vs

Day-to-day data management

- Applied data management
- Detailed strategy
- Clear rules
- Focus on immediate needs throughout the life-cycle
- Actual solutions

i.e. Fixing rules; drafting a consent form

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]




3. Why manage your data

There are many good reasons for managing your data.
Main ones include:

1. Open data movement
2. Meet funders and journals requirements
3. Legal/ethical requirements
4. For good science

Open data movement

"**Open data** is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. The goals of the open data movement are similar to those of other "open" movements such as open source, open hardware, open content and open access". Wikipedia

-  Data as public goods
-  Reproducibility
-  Further uses

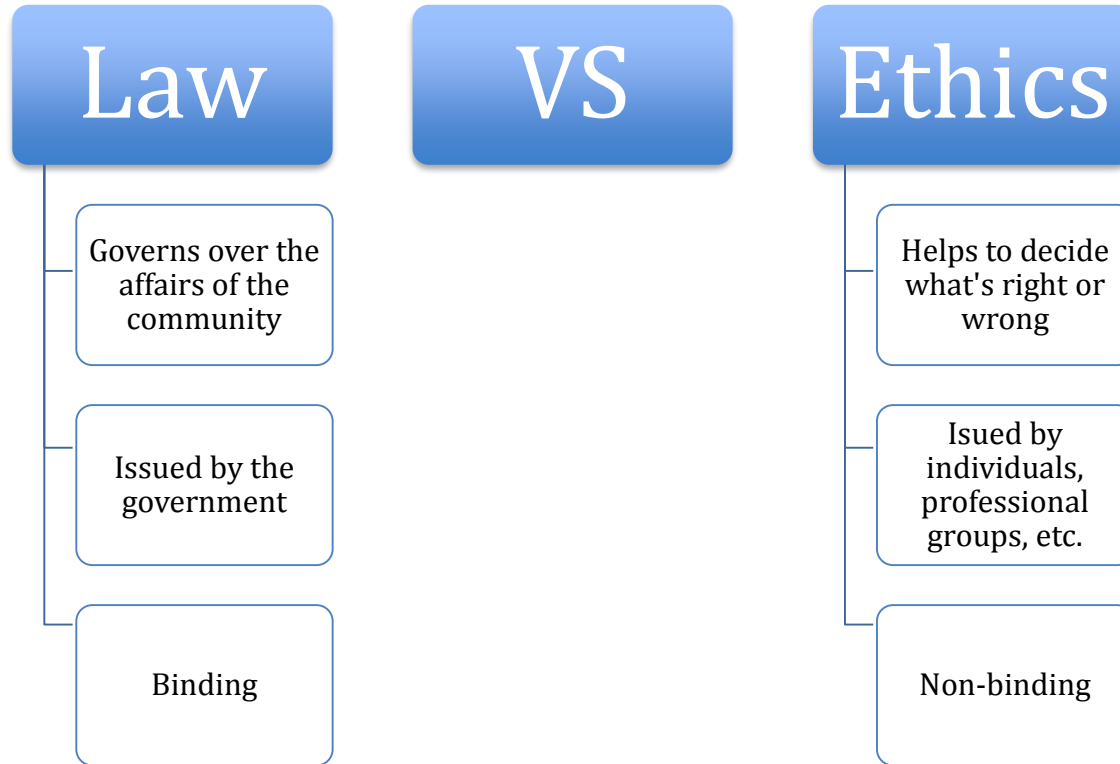
Meet funders and journals requirements

Resulting from the open data movement, more and more funders and journals require data to be made accessible.

- ➔ More and more funders require DMPs and data sharing
- ➔ More and more journals make data sharing a condition for publications

Findable 
Accessible 
Interoperable 
Reusable 

Legal and ethical requirements



Research ethics is at the heart of the research process. It consists of rules and practices that aim to protect human subjects taking part in the research, the researcher, and the team, as well as the data and possible uses that may be made out of them.

- ➔ Protection of the participants (informed consent, anonymisation, controlled access)
- ➔ Protection of the researchers (i.e. copyright, agreements)
- ➔ Protection of the data (enabling a life beyond the project, controlled access)

Researchers need to ensure that they are aware of all the relevant national and international laws that may affect their research project (i.e. data protection and intellectual property)

For good science

- It may add analytical potential (i.e. documentation)
- It increases quality of your data (consistency, documentation, etc.)
- It makes data usable by yourself and others later
- It's a precondition for the survival of your data!



And it makes research less stressful

4. How to manage data

What data do I have/will I be collecting?

It is important to question the nature of the data at the start of the research project...

- *Personal data?*
- *Sensitive data?*
- *Copyright data?*
- *Are they really necessary?*

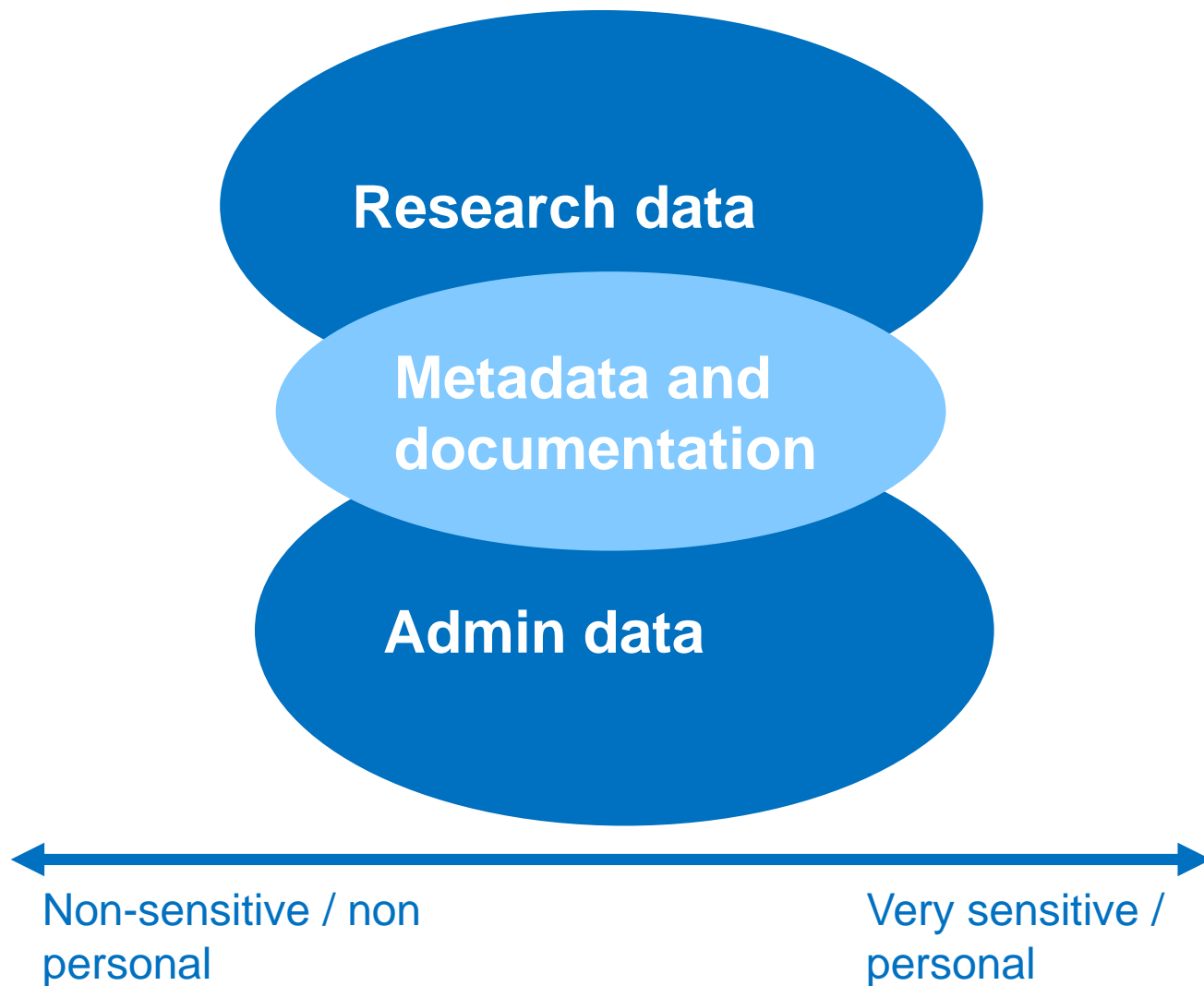
Answers to these questions directly influence the data management strategy

⚠ Sensitive data require stronger data management

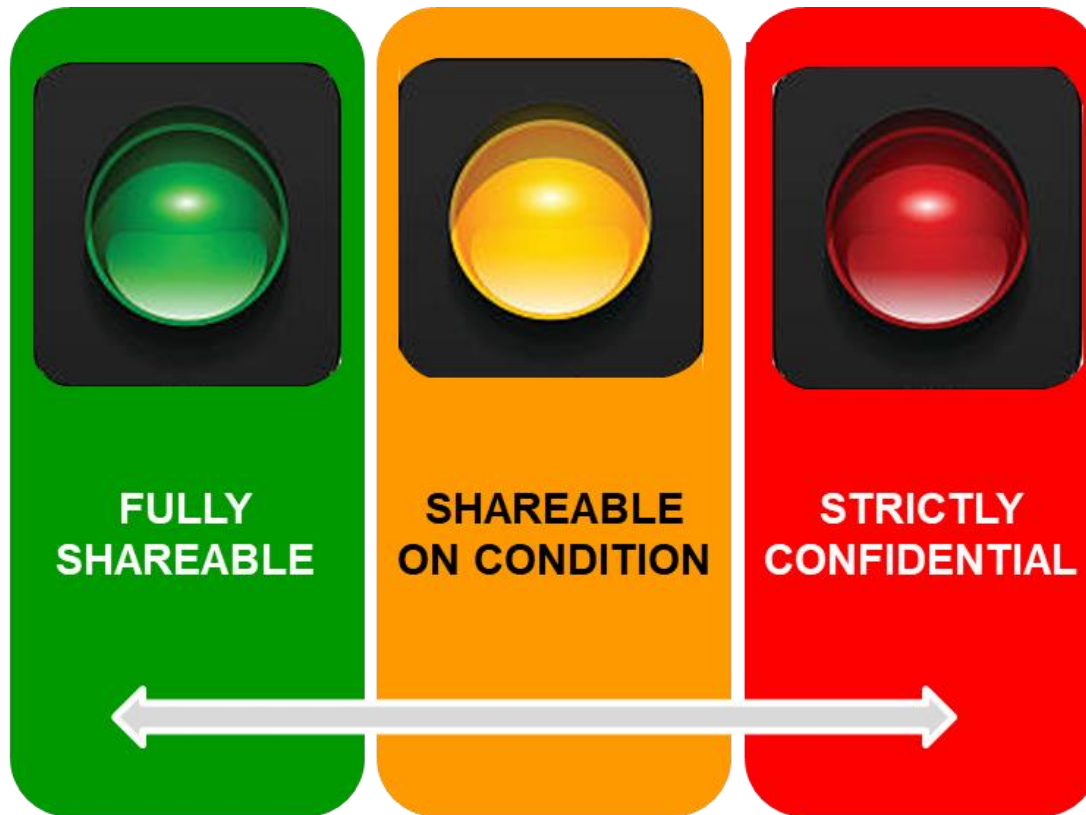
Answers to these questions also help define which data can or cannot be shared

Keyword: anticipation

➡ Saving of time at the end of the project



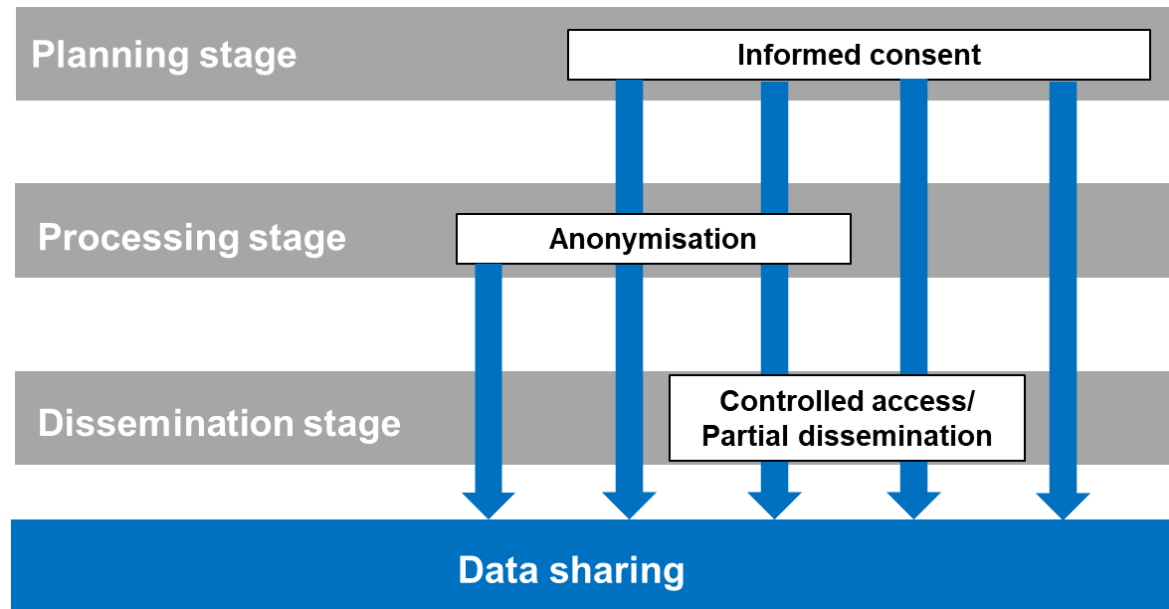
Data shareability...



Define the nature and 'shareability' of data...

- Data that are not subject to any legal and/or contractual provision can easily be shared
- Data that are subject to legal and/or contractual provisions can only be shared with adequate authorisations (consent) or following anonymisation. Ethical approval is always welcome.
- Data considered at risk or subject to special contractual provisions cannot be shared (seldom)





From planning to practice

DMPs offer a global strategy. Once funding is obtained, it is important to further develop the strategy according to the various phases of the project. In particular:

- Identify needs with respect to the project and team
- Fix rules(read-me) and apply them (file naming, versioning, etc.)
- Review rules and adapt them throughout
- Update the DMP

Remember: a good DMP is useless if not applied in practice!

From DMPs to day-to-day data management: a few examples

Planification

1. Data collection and documentation

- How to define and manage file organisation in a collective research project?
- How to fix documentation rules and limits?

2. Ethics, legal and security issues

- When ask for consent? What does 'informed' mean? What do we consent for?
- Should we go through an ethical committee?
- How to ensure security of materials during and after its treatment?
- Is anonymisation always necessary? Should we also anonymise investigators?
- How to handle multiple juridical frames as part of an international survey?

3. Data storage and preservation

- What data should be preserved beyond the project and how should it be preserved?
- Who inherit data?

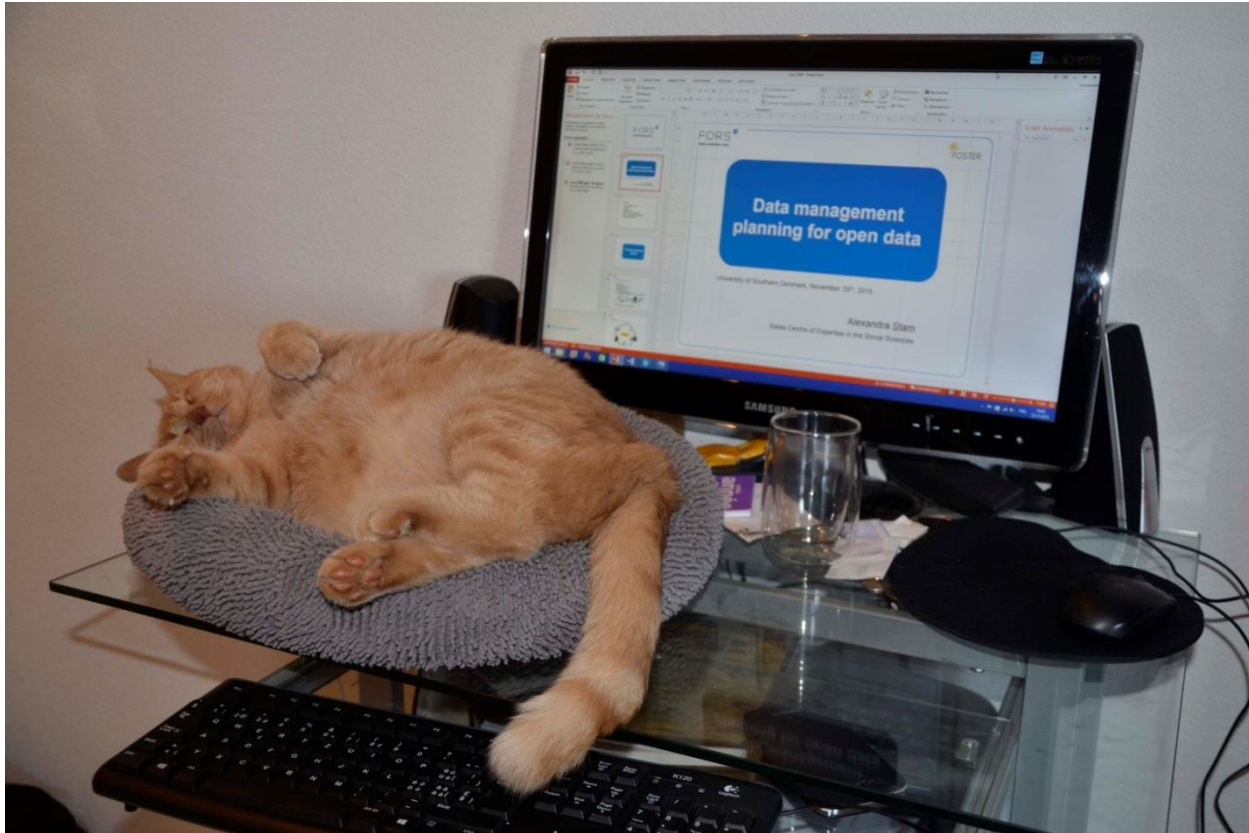
4. Data sharing and reuse

- How to ensure confidentiality in a coherent way across materials?

Recommendations

- Anticipate future as much as possible
- Consider data management as an opportunity





Thanks for your attention

alexandra.stam@fors.unil.ch

pabloandres.diazvenegas@unil.ch