

*u<sup>b</sup>*

---

*b*

**UNIVERSITÄT  
BERN**

# Data Management Planning

Module 1 Data Acquisition and Management  
CAS Applied Data Science, 23.08.2019

**Jennifer Morger and Gero Schreier, Open Science Team - University Library Bern**

[openscience@ub.unibe.ch](mailto:openscience@ub.unibe.ch), [www.unibe.ch/ub/openscience](http://www.unibe.ch/ub/openscience)

# Topics

- Introduction
- Open Science
- Data management
  - General introduction
  - File naming / Folder structure
  - Metadata & Documentation
  - Data protection
  - Storage & Backup
- Data sharing & Reuse
  - General introduction
  - Repository
  - Licenses

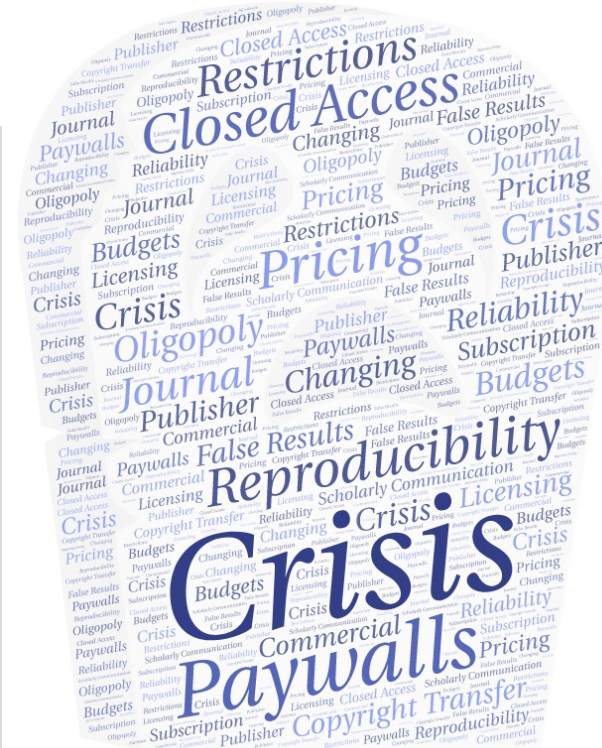
# Who are we?



Who are you?



# Issues



# Journal Crisis



Subscription prices to journals

=

skyrocketing, unsustainable prices



Access gaps

# Reproducibility Crisis

**Research**

**Restoring Study 329: efficacy and harms of paroxetine and imipramine treatment of major depression in adolescence**

BMJ 2015; 351: doi: <https://doi.org/10.1136/bmj.h4320> (Published 16 September 2015)

Cite this as: BMJ 2015;351:h4320

[Article](#) [Related content](#) [Metrics](#) [Responses](#) [Peer review](#)

Jeanne Le Nouay, research psychologist<sup>1</sup>, John M Hamlin, retired clinical assistant professor<sup>2</sup>, David Healy, & Jon Jureidini, clinical professor<sup>3</sup>, Melissa Raven, postdoctoral fellow<sup>1</sup>, Cathryn Tufanaru, research associate<sup>1</sup>, Elinor Ainsworth, staff psychiatrist<sup>1</sup>

**Author affiliations**

Correspondence to: J-J Le Nouay  
Accepted 3 August 2015

**Abstract**  
Objectives To reassess (2011), the primary objective

**Science** Home News Journals Topics Careers

**IN DEPTH** COMPUTER SCIENCE

**Artificial intelligence faces reproducibility crisis**

Matthew Hutson

• See all authors and affiliations

Science 16 Feb 2018  
10.1126/science.1251725  
DOI: 10.1126/science.1251725

**HYPOTHESIS AND THEORY ARTICLE**

**Replication, falsification, and the crisis of confidence in social psychology**

Front. Psychol., 19 May 2015 | <https://doi.org/10.3389/fpsyg.2015.00091>

**FEATURED**

**The Replication Crisis in Science**

There have been two distinct responses to the replication crisis – by instituting measures like registered reports and by making data openly available. But another group continues to remain in denial.

**Is there a reproducibility “crisis” in biomedical science? No, but there is a reproducibility problem**

Reproducibility is the key to scientific advancement. It has been claimed that we suffer from a “reproducibility crisis,” but in reality it is a chronic problem in reproducibility. Here we will look at the scope of the problem and strategies to address it.

David Gandy on June 6, 2018

**FOCUS & POLICY**

**In Medicine, the Science Has Stopped Working**

By PASCAL-EMMANUEL GODEF | November 15, 2017 4:25 PM

**Science & Environment**

**Most scientists 'can't replicate studies'**

By Tom Fildes  
Science correspondent, 'Today programme'

© 22 February 2017

**MEDICAL EDUCATION**

Tuesday, December 12, 2017 | by Greg Breining, special

**Addressing the Research Replication Crisis**

Medical schools and teaching hospitals are helping € researchers learn best practices and how to improve writing skills for research reproducibility.

**Hidden data**

The drug was widely prescribed during the swine flu outbreak in 2009.

Drug companies do not publish all their research data. This report is the result of a colossal fight for the previously hidden data into the effectiveness and side-effects of Tamiflu.

It concluded that the drug reduced the persistence of flu symptoms from seven days to 6.3 days in adults and to 5.8 days in children. But the report's authors said drugs such as paracetamol could have a similar impact.

On claims that the drug prevented complications such as pneumonia developing, Cochrane suggested the trials were so poor there was “no visible effect”.



# Reproducibility Crisis

## Is There a Crisis?

*IS THERE A REPRODUCIBILITY CRISIS?*



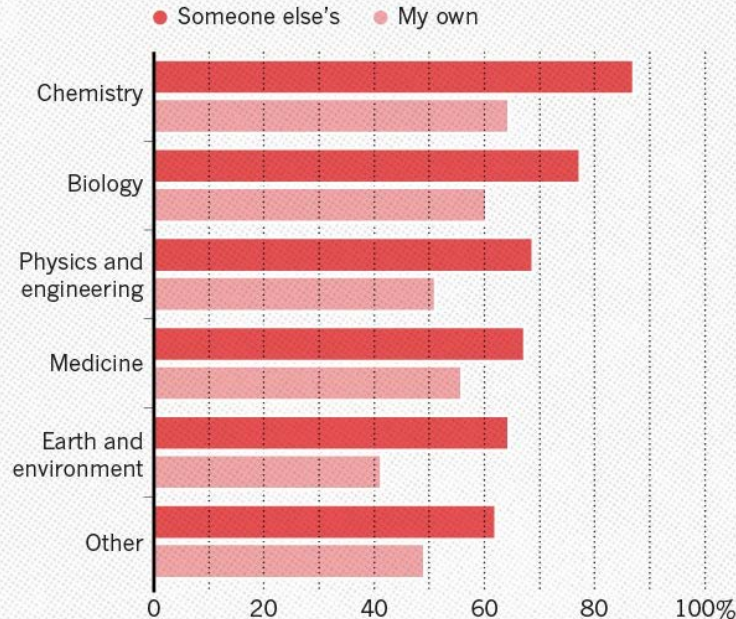
©nature

# Reproducibility Crisis

## Failed to Reproduce?

### HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



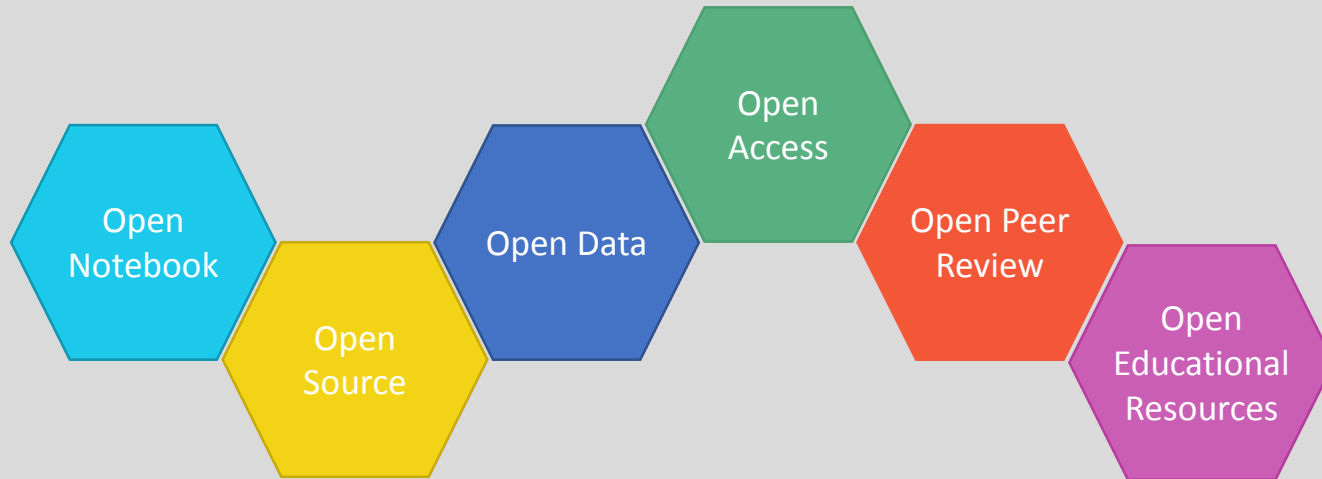
# Open Science

## Definition

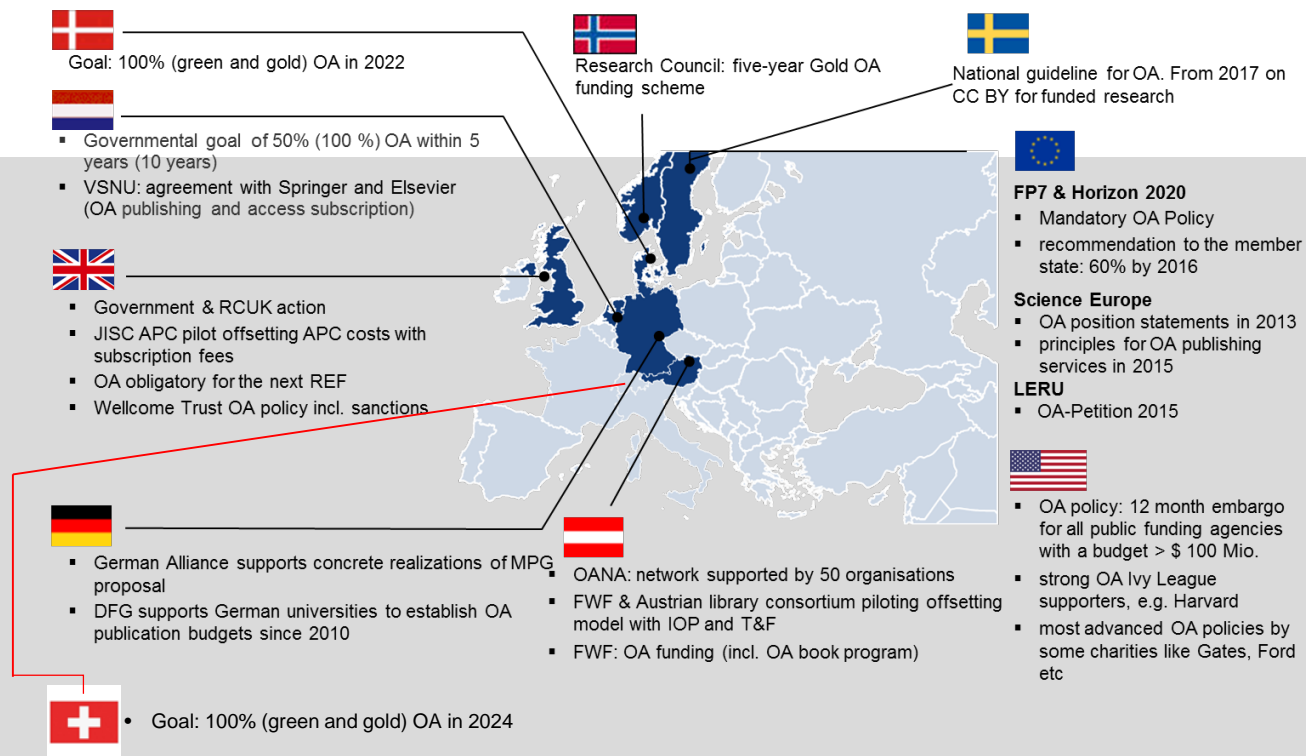
*“Open Science (OS) [is] the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.”*

Foster

# Open Science



# National Strategies – Open Access



# Open Access Policy

## University of Bern

- The University of Bern requires its researchers to deposit a full version of all peer-reviewed and published academic work and the corresponding bibliographical information in the institutional repository of the University of Bern. This makes the academic work publicly available through Open Access, provided that there are no legal obstacles.
- The University of Bern encourages its researchers to publish their research results in Open Access journals, where appropriate journals exist.

# Funder Guidelines

## Publications & Data



The results of research financed by public funds are regarded as a public good and should be published electronically so that they are immediately available without charge and can be reused by third parties. The SNSF supports the principle of free accessibility: it has adopted the aim that all publications resulting from its funding will be openly accessible as of 2020.

### The Horizon 2020 Open Access Mandate

In Horizon 2020, the European Commission (EC) requires that all peer-reviewed publications resulting from project funding are open access (OA), i.e., freely available online with no restrictions on use.





### Data Availability

**The following policy applies to all PLOS journals, unless otherwise noted.**

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction at the time of publication. When specific legal or ethical requirements prohibit public sharing of a dataset, authors must indicate how researchers may obtain access to the data.

When submitting a manuscript, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the accepted article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication.



# Data Management

## Introduction



Digitalbevaring.dk

# Data Management

## Introduction

- General introduction to data management
- Data management in practice:
  - File naming, folder structuring
  - Documentation and metadata
  - Data protection

# Data Management

## A definition

“Administrative process by which the required data is acquired, validated, stored, protected, and processed, and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users.”

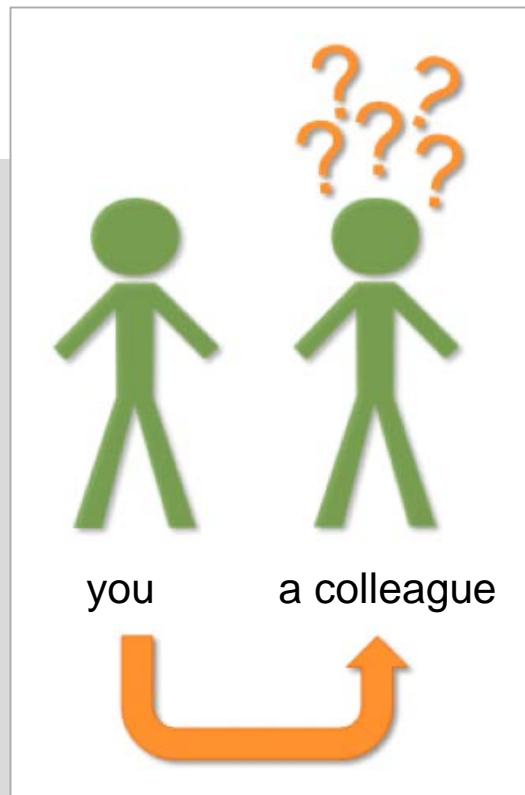
BusinessDictionary

<http://www.businessdictionary.com/definition/data-management.html>

# Data Management

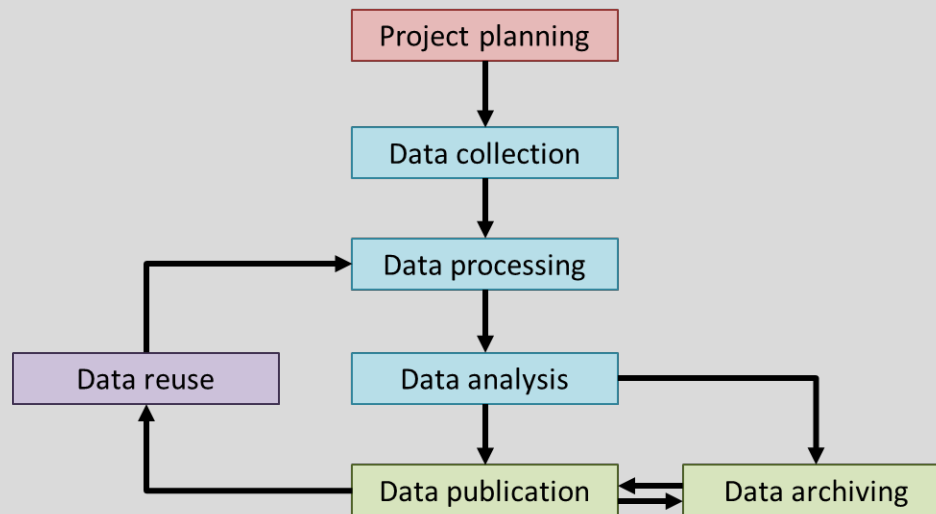
## Why?

Help others to make sense of your data – and yourself at a later point!



# Data Management

## Data Life Cycle



# Data Management DM Planning

- requirement by major research funders
- helps you to ...
  - keep track
  - address all relevant questions systematically and in advance
- DMP as a «living document»



The screenshot shows the 'mySNF' web interface for creating a Data Management Plan (DMP). At the top, there's a navigation bar with 'Inbox' and 'Tasks' tabs, and a notification for an interruption on 14.09.2019. The main content area includes instructions to complete the DMP form in the same language as the research plan, a disclaimer about the information not being part of the scientific evaluation, and a link to detailed guidelines. Below this, there's a checkbox for 'I do not submit a DMP for the following reason:'. The form is divided into four main sections, each with a dropdown arrow and a red exclamation mark icon, indicating required information:

- 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 All digital repositories I will choose are conform to the FAIR Data Principles.
  - 4.4 I will choose digital repositories maintained by a non-profit organisation.

# Data Management

## How not to: a Data Management horror story



<https://www.youtube.com/watch?v=N2zK3sAtr-4>, NYU Health Sciences Library, CC BY

# Data Management

- What experiences do you have with data management?
- Can you relate to the issues raised in the video?
- Were/are you facing special challenges with your data (e.g. volume, sensitivity ...)?



Digitalbevaring.dk



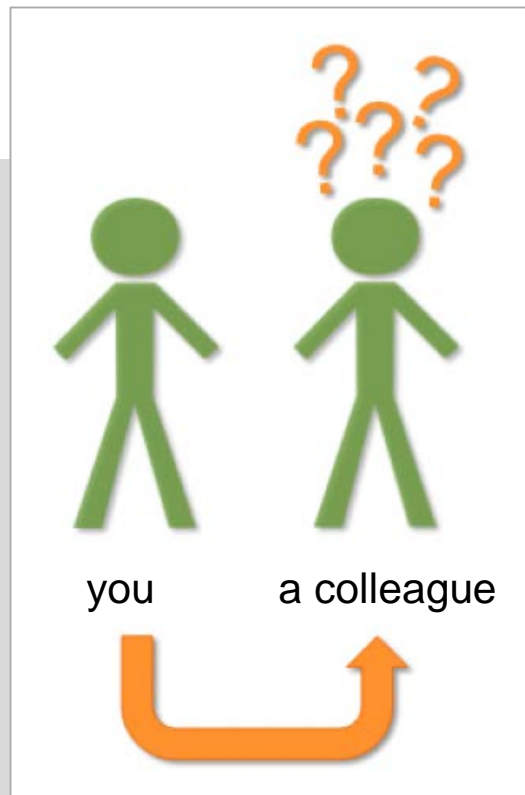
# Organization and Naming Convention



# Data Management

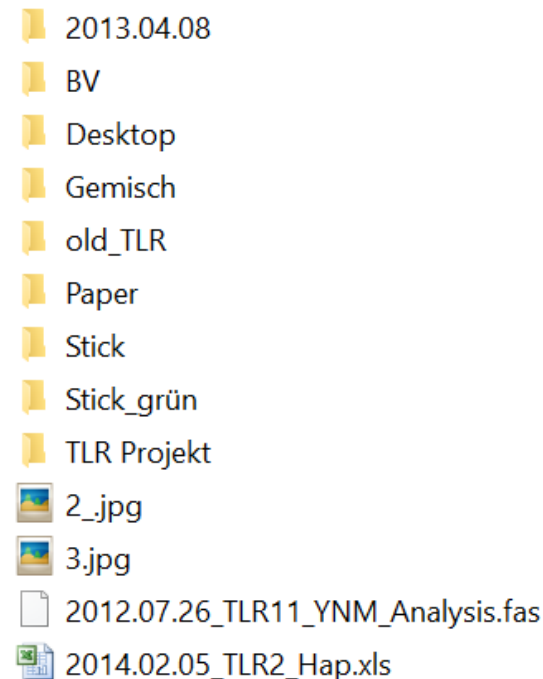
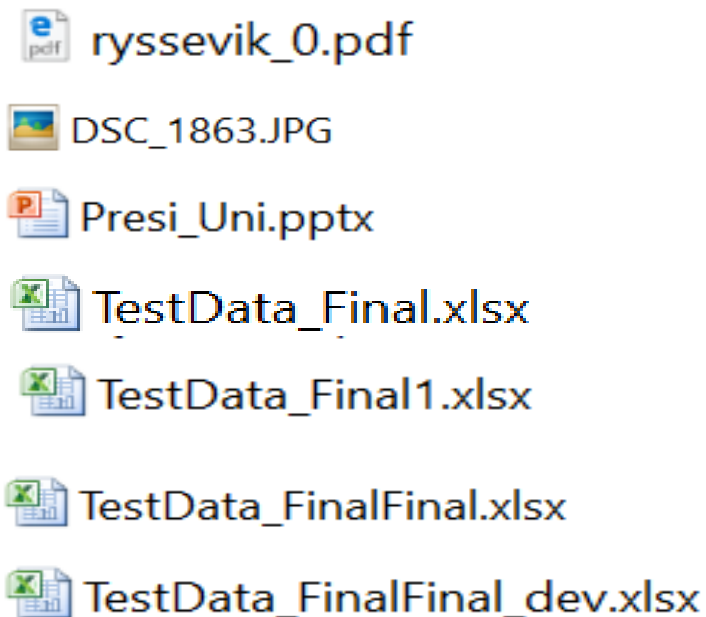
## Why?

Help others to make sense of your data – and yourself at a later point!



# Organizing your data

## How not to



# Organizing your data

## Best practices

- Be systematic and consistent
- Start early
- Balance between too much and too little
- Who will the system have to work for: You? Lab Group? Collaborators?

# Organizing your data

## Basic principles – overview

1. Directory structure
2. File naming conventions
3. File version control

# Organizing your data

## Basic principles – overview

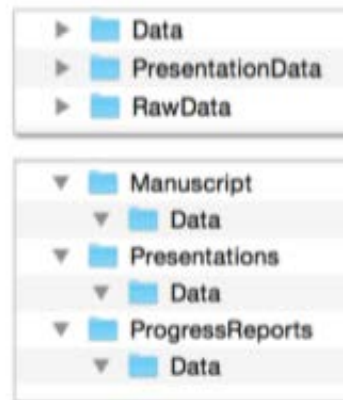
1. Directory structure
  - a. Folders and subfolders
  - b. Tagging
2. File naming conventions
3. File version control

# Structuring your data

## Folders and subfolders

- Avoid overlapping categories
- Don't let your folders get too big ("fit in one screen")
- Don't let your structure get too deep ("no more than 4 clicks")
- Use shortcuts
- keep track of your structure

### Overlapping categories

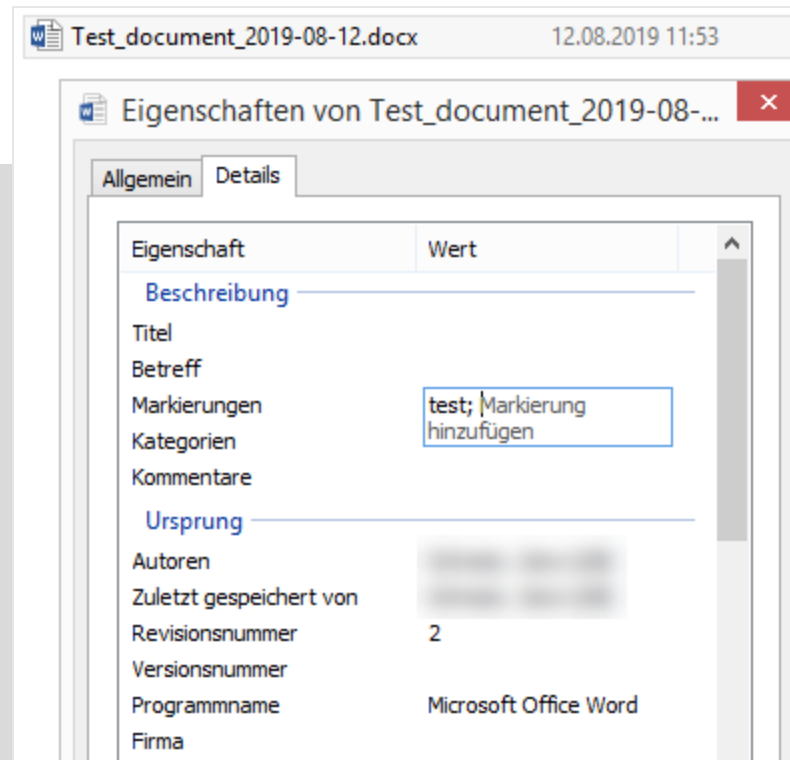


Rule of thumb:  
***"sure of the right subdirectory"***

# Structuring your data

## Tagging

- include tags for date and status (“done, pending,” etc.)
- be consistent (wording, spelling): create a master list
- one tag: max. 2 words
- tagging supported by OS vs. tagging tools (e.g. TagSpaces)





# Structuring your data

## Advantages and drawbacks of strategies

Strategy	Advantages	Drawbacks
Folders and subfolders	<ul style="list-style-type: none"><li>+ good representation of information structure</li><li>+ Similar items stored together</li></ul>	<ul style="list-style-type: none"><li>- 1 item, 1 place</li><li>- Difficult to change once set up</li></ul>
Tagging	<ul style="list-style-type: none"><li>+ 1 item, several places</li><li>+ Easier to set up and change</li></ul>	<ul style="list-style-type: none"><li>- Risk of inconsistency</li><li>- May be difficult to implement technically</li></ul>

# Structuring your data

## Tips

- Combine folders and tags
- Use tags / folders for uncharacterized files
- Use an archive folder
- Reassess your structure periodically
- Use your structure – don't collect files on your desktop ;)



<https://www.iqbginc.com/starting-records-management-program/>

# Structuring your data

## Task – 10 minutes

Please create a folder structure using the following criteria. Work in groups of 2 or 3. The folder naming is up to you.

- a. You are working on a survey project together with a colleague.
- b. The project involves data of a consumer and a stakeholder survey. Both surveys have different methodologies and questionnaires.
- c. You are working on a presentation of survey results for a team meeting and a more lengthy analysis for your superior. Sometimes your colleague sends you new file versions via email.
- d. You foresee that you will be working on the project for half a year. You will be revising questionnaires, methodologies and analyses several times and produce many new versions.

# Structuring your data

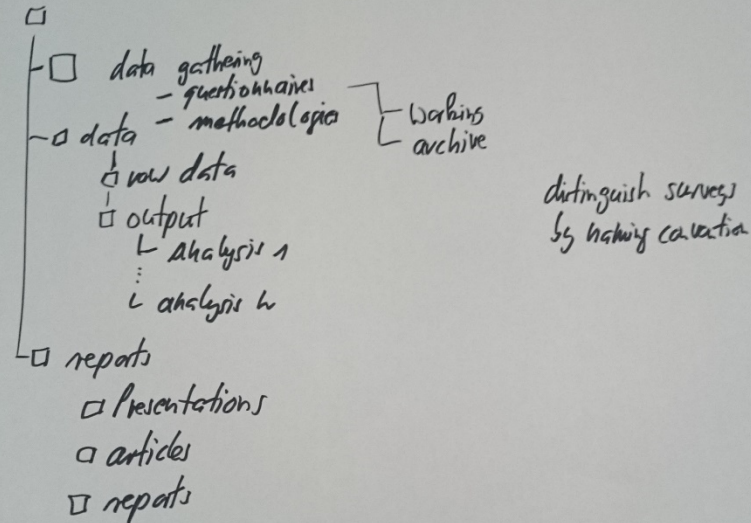
## Task – possible answer

- └─ \_ExampleProject
  - └─ 01\_Data
    - └─ ConsumerSurvey
    - └─ StakeholderSurvey
  - └─ 02\_Documentation
    - └─ Methodology
      - └─ Method\_ConsumerSurvey
      - └─ Method\_StakeholderSurvey
    - └─ Questionnaires
      - └─ QuestionnaireConsumerSurvey
      - └─ QuestionnaireStakeholderSurvey
  - └─ 03\_Analysis
    - └─ AnalysisEmails
  - └─ 04\_Presentation
    - └─ PresentationEmails
  - └─ 05\_Archive
    - └─ AnalysisArch
    - └─ DocumentationArch
    - └─ PresentationArch

based on [https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure, example survey data](https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure,example%20survey%20data)

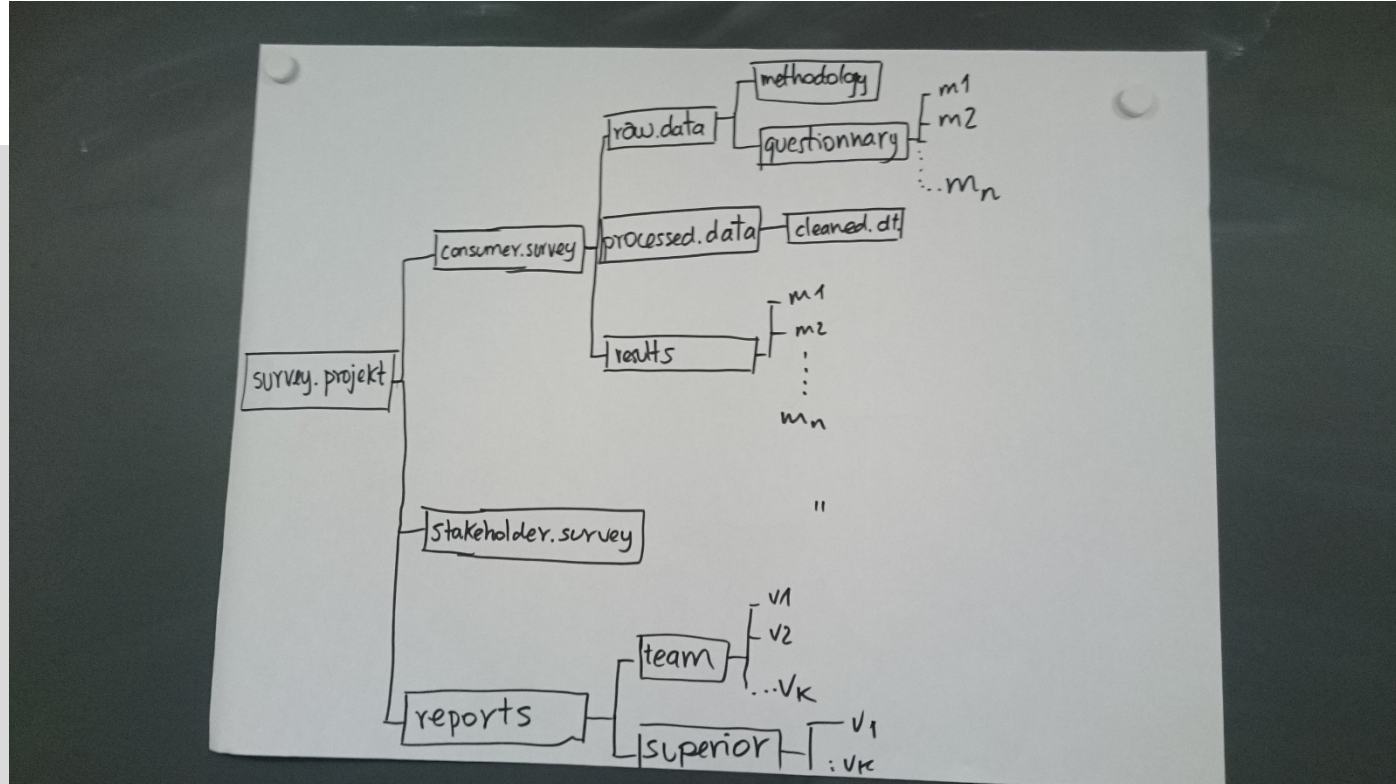
# Structuring your data

## Group 1



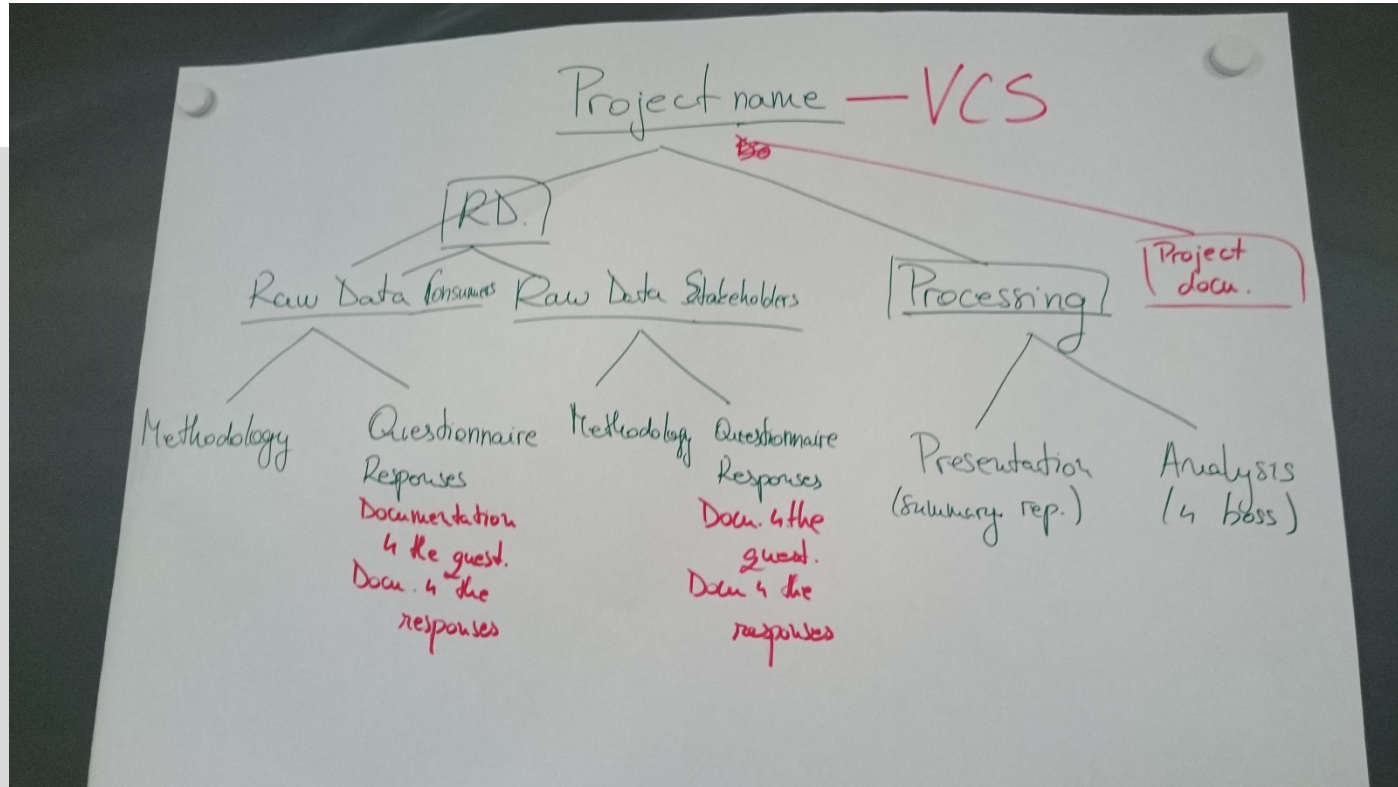
# Structuring your data

## Group 2



# Structuring your data

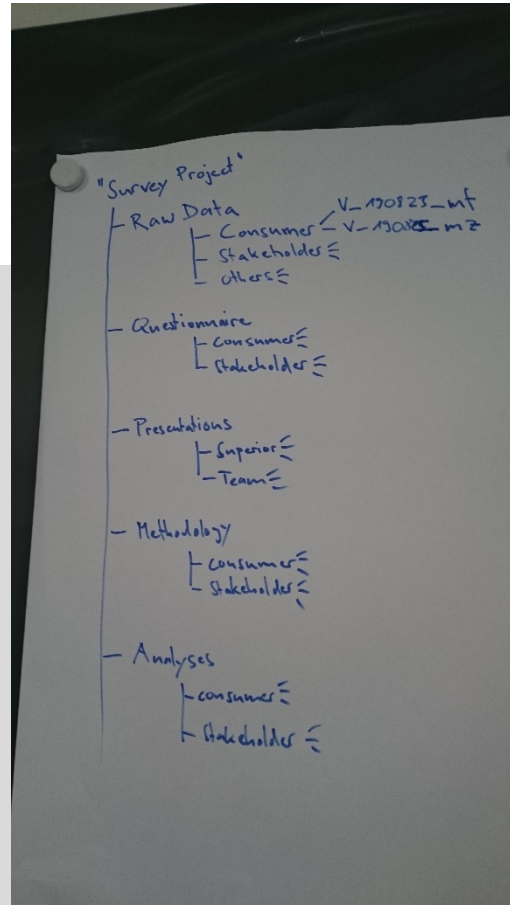
## Group 3





# Structuring your data

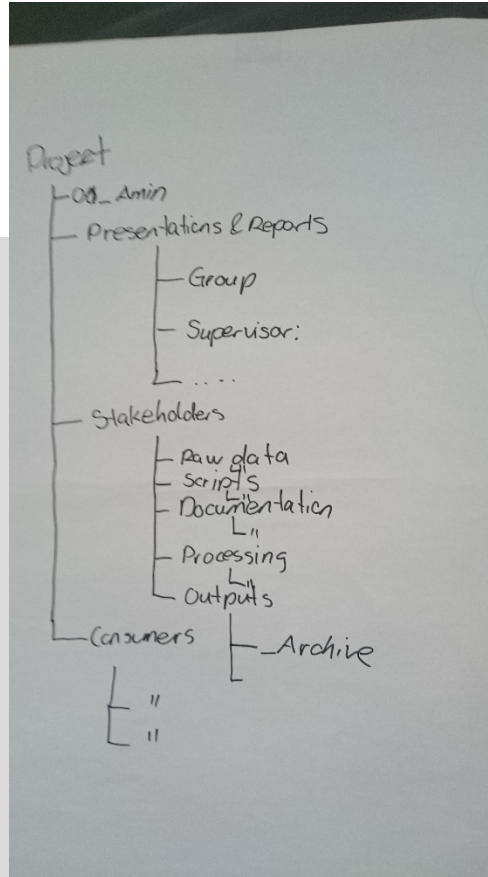
## Group 4





# Structuring your data

## Group 5



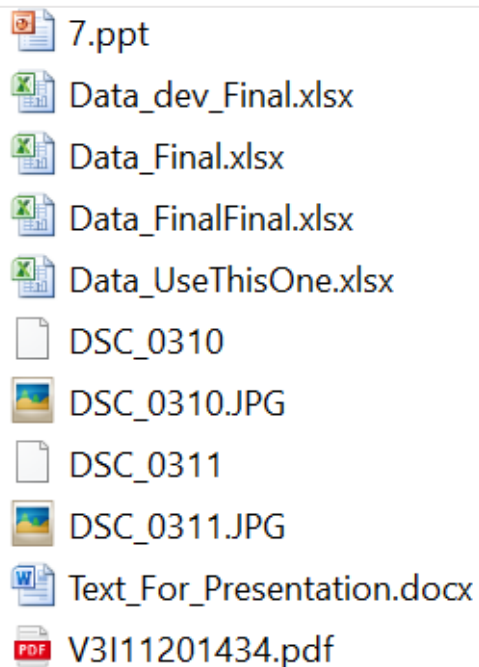
# Organizing your data

## Basic principles – overview

1. Directory structure
2. File naming conventions
3. File version control

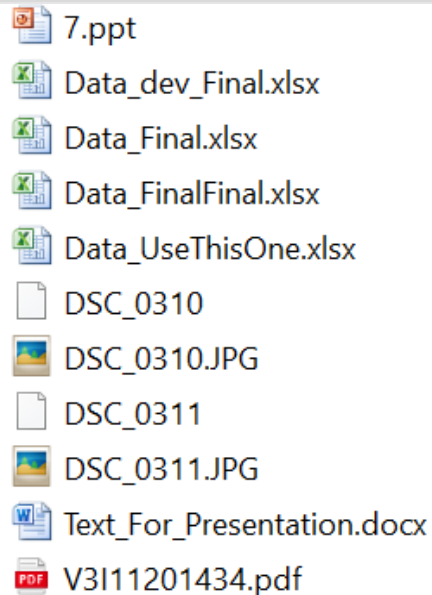
# Organizing your data

## File Naming



# Organizing your data

## File Naming Task – 10 minutes



7.ppt  
Data\_dev\_Final.xlsx  
Data\_Final.xlsx  
Data\_FinalFinal.xlsx  
Data\_UseThisOne.xlsx  
DSC\_0310  
DSC\_0310.JPG  
DSC\_0311  
DSC\_0311.JPG  
Text\_For\_Presentation.docx  
V3I11201434.pdf

Work in groups of 2-3. Develop a naming convention for your files.

- What are some general considerations when choosing a file name?
- Which parts should it contain?
- Determine the sequence of the components and explain why.
- Where would you document your naming convention?

# File Naming Conventions

## Tips

### Include

- name or initials
- Date
- Version number
- Unique identifier (e.g. project number)
- Project name

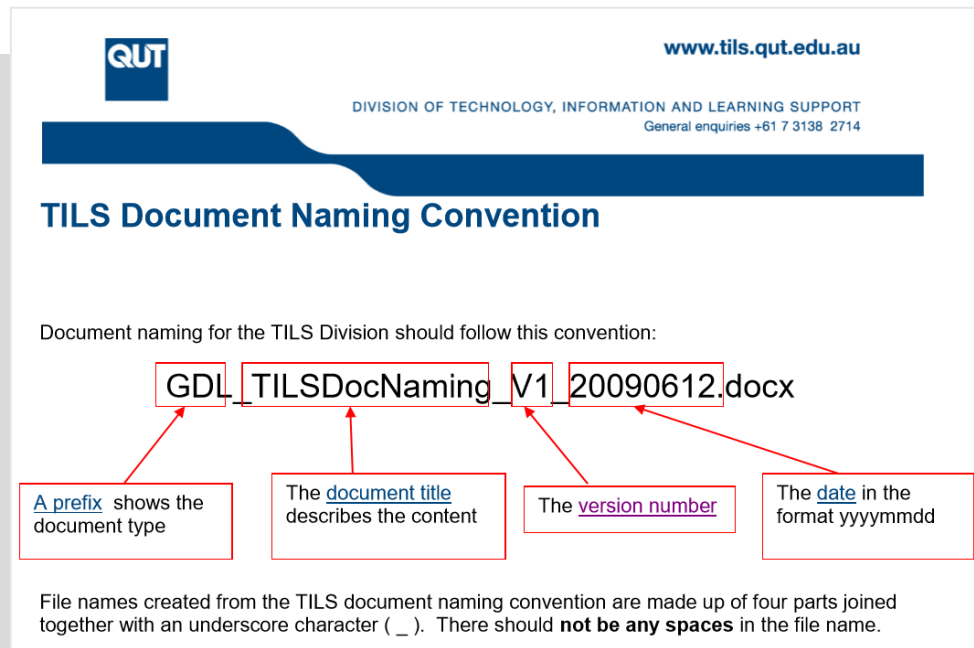
### Avoid

- spaces (use hyphens or underscores)
- special characters (e.g. & ( ) [ ] “ ”)

Document your convention (readme-file)

# File Naming Conventions

## Example: TILS



The slide features a header with the QUT logo, the website [www.tils.qut.edu.au](http://www.tils.qut.edu.au), and contact information for the Division of Technology, Information and Learning Support. The main title is 'TILS Document Naming Convention'. Below this, a paragraph states that document naming should follow a specific convention. A diagram shows the file name 'GDL\_TILSDocNaming\_V1\_20090612.docx' with red boxes around each part and arrows pointing to explanatory text boxes. The parts are: 'GDL' (document type), 'TILSDocNaming' (document title), 'V1' (version number), and '20090612' (date in yyyymmdd format). A final paragraph explains that file names are composed of four parts joined by underscores and should not contain spaces.

**QUT** [www.tils.qut.edu.au](http://www.tils.qut.edu.au)  
DIVISION OF TECHNOLOGY, INFORMATION AND LEARNING SUPPORT  
General enquiries +61 7 3138 2714

### TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:

**GDL\_TILSDocNaming\_V1\_20090612.docx**

- A prefix shows the document type
- The document title describes the content
- The version number
- The date in the format yyyymmdd

File names created from the TILS document naming convention are made up of four parts joined together with an underscore character ( \_ ). There should **not be any spaces** in the file name.

[https://www.data.cam.ac.uk/files/gdl\\_tilsdocnaming\\_v1\\_20090612.pdf](https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf)

# File Naming Conventions

## Example: TILS

GDL\_TILSDocNaming\_V1\_20090612.docx

PRE\_LibDatabaseMgmt\_V1\_20090124.ppt

Prefix	Meaning
AGD	Agenda
AGR	Agreement
GDL	Guideline
MEM	Memorandum
MIN	Minutes and Notes
PRE	Presentation
PRO	Procedure
PRP	Proposal
REP	Report
TEM	Template

# File Naming Conventions

## Example: TILS

### **GDL\_TILSDocNaming\_V1\_20090612.docx**

- Version 1 of the TILS Document Naming guidelines prepared on the 12<sup>th</sup> of June 2009

### **PRE\_LibDatabaseMgmt\_V1\_20090124.ppt**

- A powerpoint presentation about database management prepared by the Library on the 24<sup>th</sup> of January 2009



# File Naming Conventions

Example: Stanford Best Practice

FR3S.140623.129C.2653.W.JPG

# File Naming

## Example: Stanford Best Practice

### Info tracked & the convention used

The researchers wanted to track several things about the tiles:

1. **Study site.** Indicated by the name, ex. FR3, FR7, FR9.
2. **Depth of the water.** Indicated by S (shallow), M (middle), or D (deep).
3. **Date.** Indicated by YYMMDD.
4. **Tile number.** Indicated on the tile.
5. **Tile treatment.** Indicated by C (caged) or U (uncaged).
6. **Number assigned to photo by camera.**
7. **Whether the post-removal photo was of the entire tile or a tile section.**  
Indicated by W (whole area), A (upper right), B (lower right), C (lower left), or D (upper left).

Example: FR3S.140623.129C.2653.W.JPG

This was image 2653 of whole, uncovered tile 129 from study site 3 in shallow water, taken on June 23, 2014.

# File Naming Conventions

## Renaming Files - Tools

Batch/bulk renaming tools e.g.:

- [Ant Renamer](#) (Windows)
- [Renamer 5](#) (Mac)
- [GNOME Commander](#) (Linux)

Tips:

- make sure the software doesn't change the file format
- keep track of original file names

# Organizing your data

## Basic principles – overview

1. Directory structure
2. File naming conventions
3. File version control

# Organizing your data

## Version Control

**GDL\_TILSDocNaming\_V1\_20090612.docx**

- Version 1 of the TILS Document Naming guidelines prepared on the 12<sup>th</sup> of June 2009

- Revert to previous versions
- Find out what is different between two versions
- Find out what has changed in a specific time period
- Manage multiple versions
- Work with multiple people on the same files
- Transparency and integrity

# Version Control

## How to – simple

### Versioning in file names

- Ordinal numbers for major and decimals for minor changes → file\_V1-2
- Use dates to distinguish between versions or add to version numbers → file\_V1-2\_2019-08-13
- Use the label “final” – but only once 😊 → file\_V1-2\_2019-08-13  
file\_V1-3\_final

# Version Control

## How to – simple

### Tips

- Use an archive folder
- Delete versions you don't need anymore
- document your convention (readme-file)

# Version Control

## How to – elaborate

### Versioning systems

- [Git](#)
- [Mercurial](#)
- [Bazaar](#)
- [Darcs](#)

... for text or table files

- automated versioning in cloud applications (Office 365, GoogleDocs)
- other software, e.g. Word 2016-19, [Simul](#)



# Organizing your data

Any Questions?

# Metadata & Documentation



# Metadata

## What is it?

- Commonly described as “data about data”
- Structured information that conforms to standards
- Make data findable, reusable and citable
- For a dataset e.g.: Title, Creator, Description, Format, Rights,...
- Metadata schemas: e.g. Dublin Core, Data Cite

# Metadata

## Where is it?

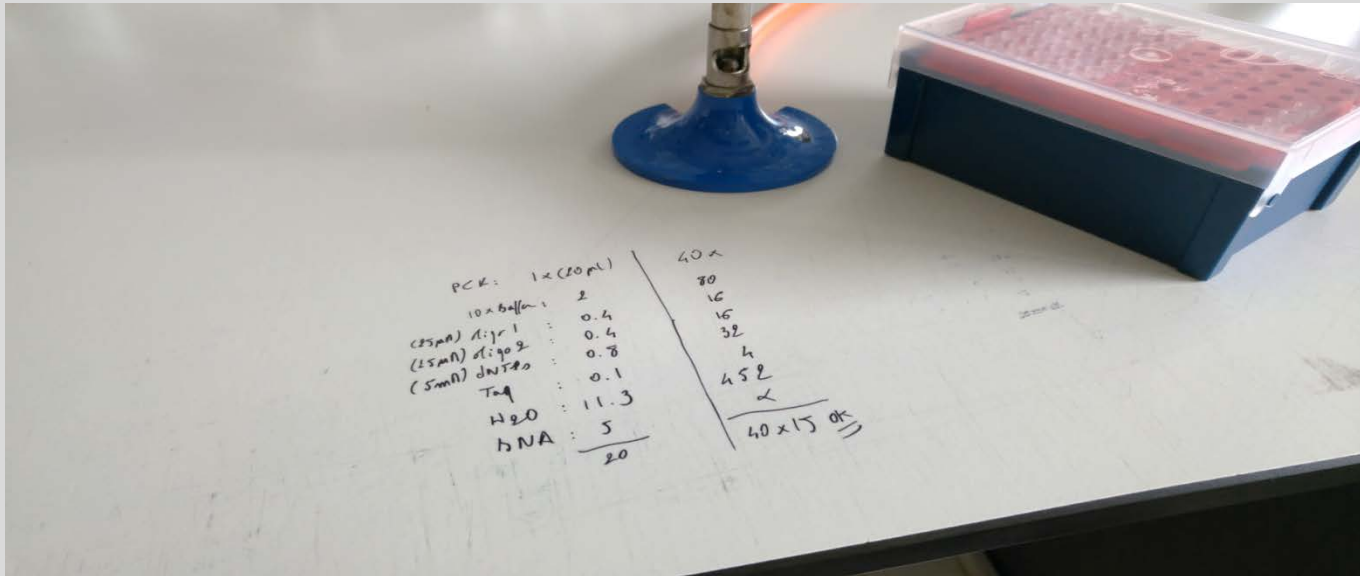
- In your raw/processed data
- Manual input
- Automatic input (from lab tools)
- In your documentation [Readme.txt, log files, Info files, Submitting interface]
- File & folder name
- In a data repository [where you will deposit your data]

## 61

[illegible]

# Documentation

## How do I document?



# Documentation

## How do I document?

- Document your data *throughout* your research – do not wait till the end!
- A few tips for continuous documentation:
  - Try a generic note-taking software (e.g. SCINote, OneNote, Evernote,...)
  - Use an electronic lab notebook for structured documenting
  - If you work with scripting languages, such as R or Python, take a look at [Jupyter Notebook](#).

# ELN yes, but which one?

## More information

- Practical [guidelines](#) on how to introduce ELM and LIMS in an academic research laboratory by the DLCM Team.
- Lists of available products:
  - ELM list by the University of Harvard: [comparative table](#)
  - DLCM list of [ELNs](#) and [LIMS](#)
  - [Website](#) with ELM list and additional information by the Gurdon Institut (University of Cambridge)

Features	Specifications							
	Benchling	Biovia	Confluence	Doccollab	ECL	ELOG	Evernote	Exemp
<b>Interactivity</b>								
Intuitive Interface Design	✓	No response received	✓	✓	No response received	✓	No response received	No response received
Auto Metadata Harvest	✓	No response received	✗	✓	No response received	✗	No response received	No response received
Search functions can search across file formats and beyond types	✓	✓	✓	✓	No response received	✓	✓	✓
Ability to manipulate files and images	✓	No response received	✓	✓	No response received	✓	No response received	✓
Support for multiple open windows	✓	No response received	✓	✓	No response received	✓	No response received	✓
Ability to link out	✗	No response received	✓	✓	✓	✓	✓	✓
<b>Support for Researcher Documentation</b>								
Hyperlink support	✓	No response received	✓	✓	✓	✓	✓	✓
Metadata Creation Prompts	✓	No response received	✗	✓	No response received	✗	✗	No response received
Rights Management (licensing)	✓	No response received	✓	✓	✓	✓	No response received	No response received
Protocol Integration	✓	✓	✓	✓	No response received	✓	✓	✓
<b>Adaptability to Lab workflows</b>								
Accounts/Permissions Levels	✓	No response received	✓	✓	✓	✓	✓	✓
Internal Data Sharing	✓	✓	✓	✓	No response received	✓	No response received	✓
Adaptable to a Variety of Workflows	✓	No response received	✓	✓	No response received	✓	No response received	✓
Compatibility with authoring tools	✓	No response received	✓	✓	No response received	✗	No response received	No response received
Windows Compatible	✓	No response received	✓	✓	✓	✓	✓	✓
Macintosh Compatible	✓	✓	✓	✓	✓	✓	✓	✓
Linux Compatible	✓	✗	✓	✓	No response received	✓	No response received	✓
Android Compatible	✓	✓	✓	✓	No response received	✓	✓	✓
iOS Compatible	✓	✓	✓	✓	No response received	✓	✓	✓
<b>Storage</b>								
Cloud Storage	✓	No response received	✗	✓	No response received	✓	No response received	No response received
Local Storage	✗	No response received	✓	✗	No response received	✓	No response received	No response received
Hybrid (cloud/local) Storage	✗	No response received	✗	✗	No response received	✗	No response received	No response received
Versioning	✓	✓	✓	✓	No response received	✓	No response received	✓
File Redundancy	✓	No response received	✓	✓	No response received	✓	No response received	No response received
Creates stable URLs or persistent identifiers for entities	✓	No response received	✓	✓	No response received	✓	No response received	No response received
Can unregistered users access the data found at persistent links?	✓	No response received	✓	✗	No response received	✗	No response received	No response received
Response Time (s)	✓	No response received	✓	✓	No response received	✓	✓	No response received



# Documentation

## README files

- Describe the files and folders in a project.
- Primarily aimed at an external audience and your future self
- Write as a plain text file
- Use standards
- README Template:

<https://data.research.cornell.edu/content/readme>

This DATSETNAMEreadme.txt file was generated on [YYYYMMDD] by [Name]

-----  
GENERAL INFORMATION  
-----

1. Title of Dataset

2. Author Information

Principal Investigator Contact Information

Name:  
Institution:  
Address:  
Email:

Associate or Co-investigator Contact Information

Name:  
Institution:  
Address:  
Email:

Alternate Contact Information

Name:  
Institution:  
Address:  
Email:

3. Date of data collection (single date, range, approximate date) <suggested format YYYYMMDD>

# Legal Frameworks

For working with data



# Legal Frameworks

## Please note



We are no legal experts.

This is meant only as an orientation.

For more detailed questions and for advice, please consult a lawyer or data protection officer!

# Legal Frameworks

## Overview

- Legal frameworks and definitions
- Before processing data:
  - Which data?
  - Informed consent
- Data processing: how to?
  - Anonymisation, pseudonymisation

# Legal Frameworks

## For working with data



- Federal Act on Data Protection, Switzerland (1992/2019, under revision)
- Data Protection Regulation, Bern (1986/2013)
- General Data Protection Regulation, EU (GDPR, 2018)

# Legal Frameworks

## For research involving data

### Datenschutz und Forschung im Allgemeinen

<https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/statistik--register-und-forschung/forschung/datenschutz-und-forschung-im-allgemeinen.html>

### Federal Act on Research involving Human Beings

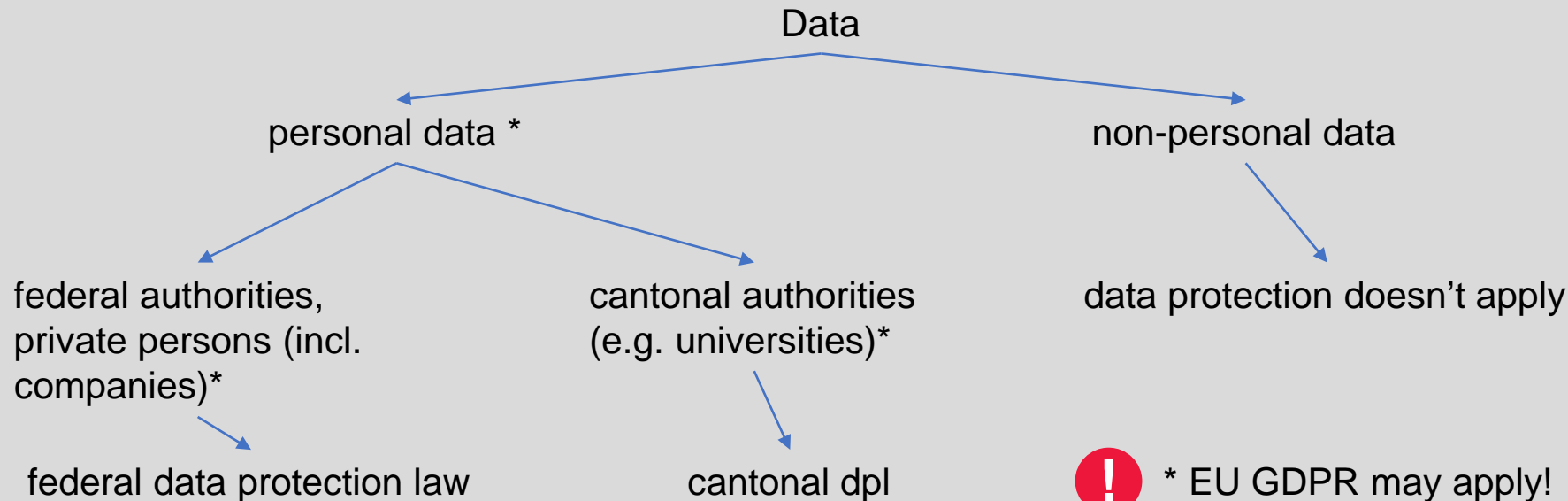
<https://www.admin.ch/opc/en/classified-compilation/20061313/index.html>



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

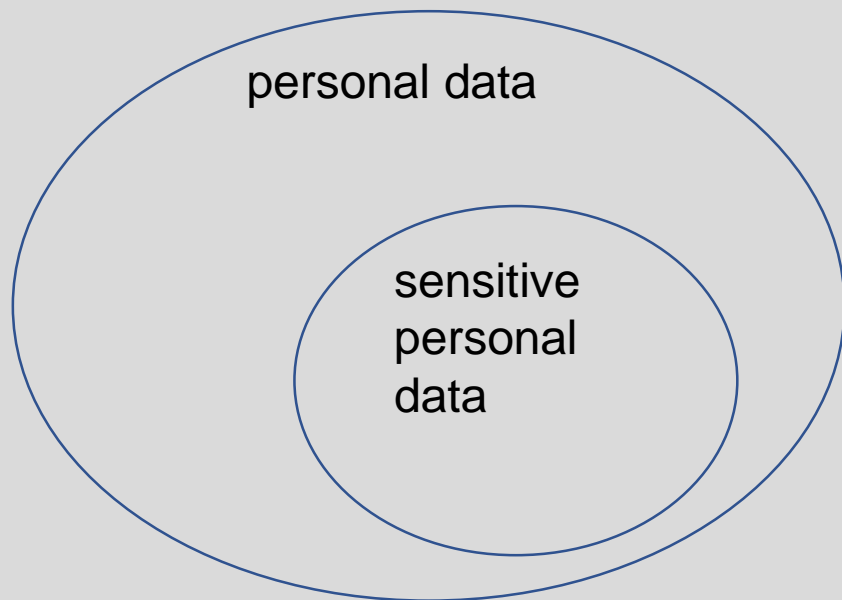
# DPR Switzerland

## Working with data



# DPR Switzerland

## Some definitions



### **Personal Data**

- Any data that can be related to an identifiable or identified person

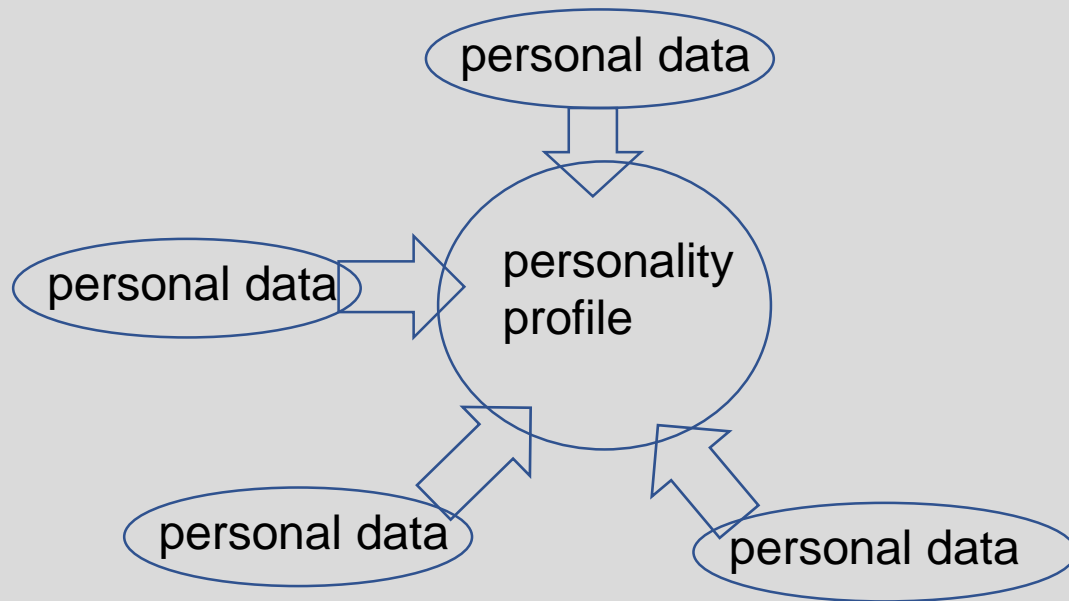
### **Sensitive Personal Data**

- religious, ideological, political or trade union-related views or activities
- health, intimate sphere, race
- social security measures
- administrative or criminal proceedings and sanctions



# DPR Switzerland

## Some definitions



### Personality Profile

- a collection of data that permits an assessment of essential characteristics of the personality of a natural person

FADP, Art. 3d

# DPR Switzerland

## Some definitions

### **Data processing**

= anything (!) that can be done with data:

collecting, storing, archiving, analysing (or other forms of use), publishing, deleting ...

(FADP, Art. 3e)

# Legal Frameworks

## Overview

- Legal frameworks and definitions
- Before processing data:
  - Which data?
  - Informed consent
- Data processing: how to?
  - Anonymisation, pseudonymisation

# DPR Switzerland

## Which Data?

- Collect only data needed to the purpose of your study (Principle of Proportionality)
- Data may only be processed to fulfill the purpose indicated when it was collected
- Data must be accurate
- Data processing must not aim at identifying a person or making a person identifiable

# DPR Switzerland

## Informed Consent

### Data subjects

- must be informed about planned data processing and their rights in advance.
- must consent to the processing of their data.
- have the right to object to processing their data.
- Tip: keep it short and simple (as far as possible)!

# Legal Frameworks

## Overview

- Legal frameworks and definitions
- Before processing data:
  - Which data?
  - Informed consent
- Data processing: how to?
  - Anonymisation, pseudonymisation

# DPR Switzerland

## Anonymisation

- Ensure that data cannot be related to a specific person, or can only be assigned with extraordinary effort
- Legal obligation!
- The data must be anonymised as quickly as possible
- The research results must be published in anonymous form
- Data protection regulations do not apply to anonymized data

# DPR Switzerland

## Anonymisation - Example

### Before anonymization

Name	Age	Sex	Income	postcode
Martin Müller	51	m	79'000	3001
Andrea Sommer	21	f	55'000	3013
Dominik Fischer	44	m	102'000	3012
Arnold Furrer	65	m	40'000	3001
Simone Meier	38	f	67'000	3011


### After anonymization

Name	Age	Sex	Income	postcode
*	41-60	m	79'000	30**
*	21-40	f	55'000	30**
*	41-60	m	102'000	30**
*	*	*	*	30**
*	21-40	f	67'000	30**



# DPR Switzerland

## Pseudonymisation

- Identifying data is replaced by an identifier or pseudonym
  - Key allows mapping of identifiers to data subjects
  - Key must be kept
    - separate from data
    - securely, encrypted
  - Must only be used if anonymization is not possible
-  Data protection regulations apply to pseudonymized data (≠ anonymisation)

# DPR Switzerland


## Pseudonymisation - example

### Before pseudonymization

Name	Age	Sex	Income	postcode
Martin Müller	51	m	79'000	3001
Andrea Sommer	20	f	55'000	3013
Dominik Fischer	44	m	102'000	3012
Arnold Furrer	75	m	40'000	3001

### After pseudonymization

Identifier	Age	Sex	Income	postcode
1	51	m	79'000	3001
2	20	f	55'000	3013
3	44	m	102'000	3012
4	75	m	40'000	3001



Name	Identifier
Martin Müller	1
Andrea Sommer	2
Dominik Fischer	3
Arnold Furrer	4

# DPR beyond Switzerland

## Europe - GDPR

### General Data Protection Regulation (2018)

- Lawful, fair, transparent
- Consent form easy to understand
- Re-purposing of data requires informed consent of data subject
- Transferring data outside EU requires that target countries provide similar protection

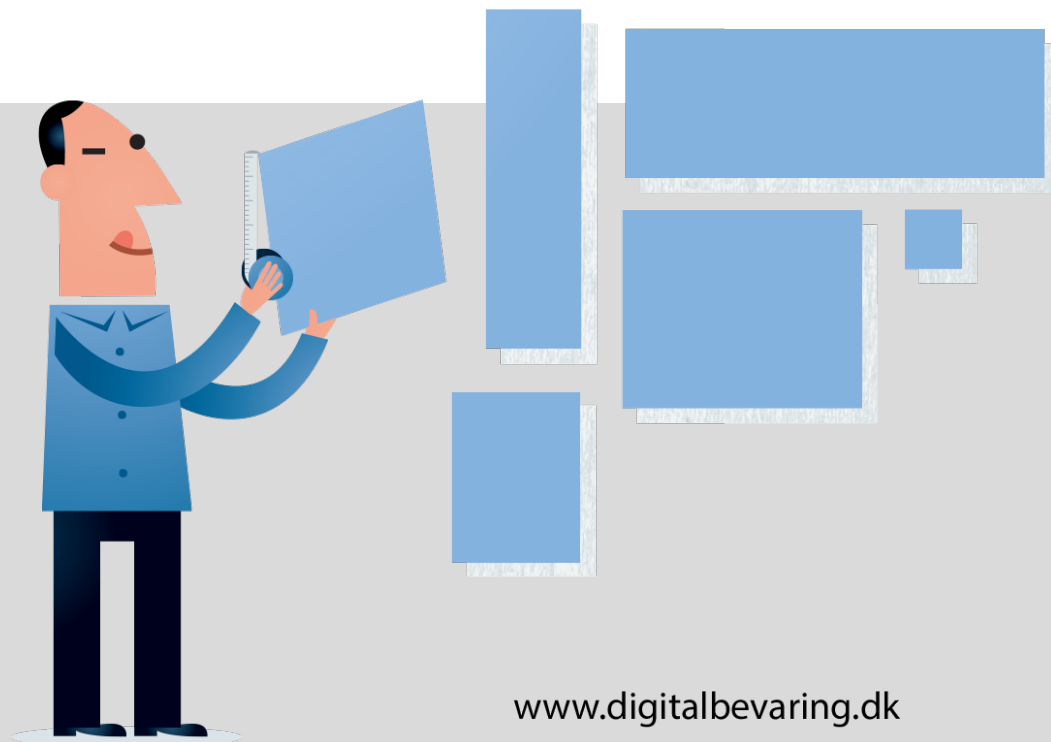


# Data protection

## Where to get more information

- [Data protection office of the Canton of Berne](#)
- [Bern University Legal Services Office](#)
- Legal texts:
  - [Federal Act on Data Protection \(1992/2019\)](#)
  - [Data Protection Act, Canton of Berne \(1986/2013\)](#)
  - [EU, GDPR \(2018\)](#)
- [datenrecht.ch](#): website on legal regulations around data (German)
- [Open Science @ UniBe](#): basic information around handling sensitive data
- and many more ...

# Data Storage & Back-up



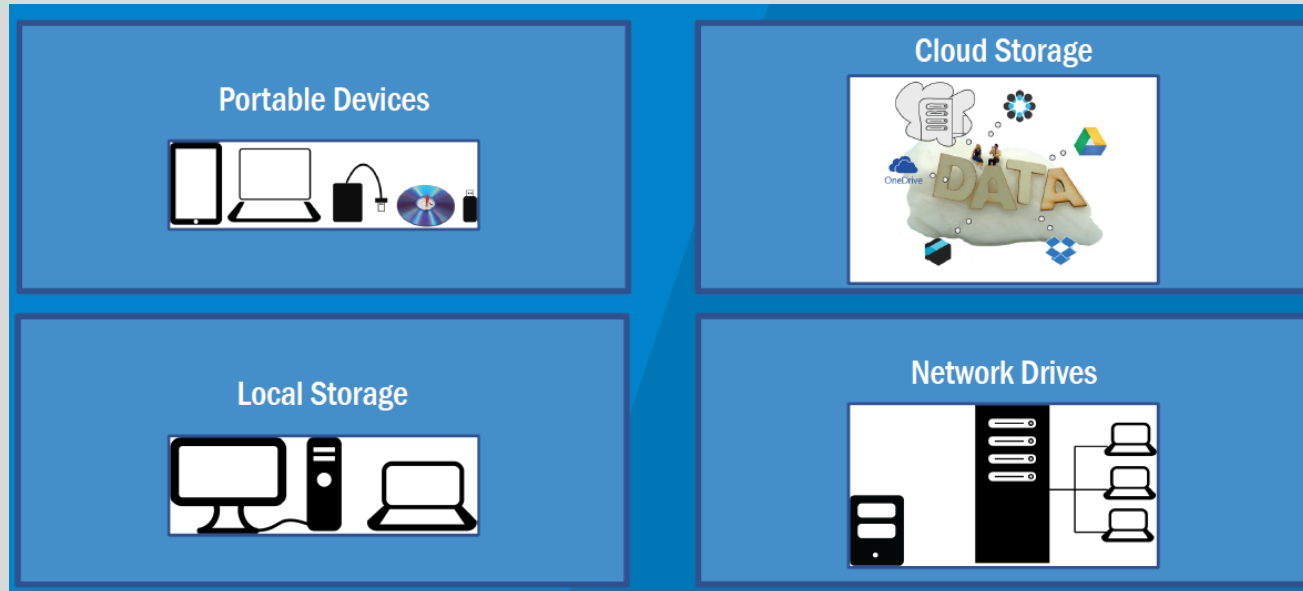
[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

# Data Storage

## Check your Needs

- How much space do you need?
- Who needs access to the data and on what level?
- Do you need extra protection for sensitive data?

Discuss with your Neighbour the advantages or disadvantages of:



# Data storage

## Storage solutions – Portable Devices

Portable Devices	
<ul style="list-style-type: none"><li>+ transport</li><li>+ exchange</li><li>+ secure</li><li>+ cheap</li></ul>	<ul style="list-style-type: none"><li>- loss / theft</li><li>- break / damage</li><li>- limited space / unflexible</li><li>- degradation</li><li>- exchange +/- cost / manage manually</li><li>- technology shifts / incompatibility</li></ul>
Local Storage	
<ul style="list-style-type: none"><li>+ </li></ul>	<ul style="list-style-type: none"><li>+ </li></ul>



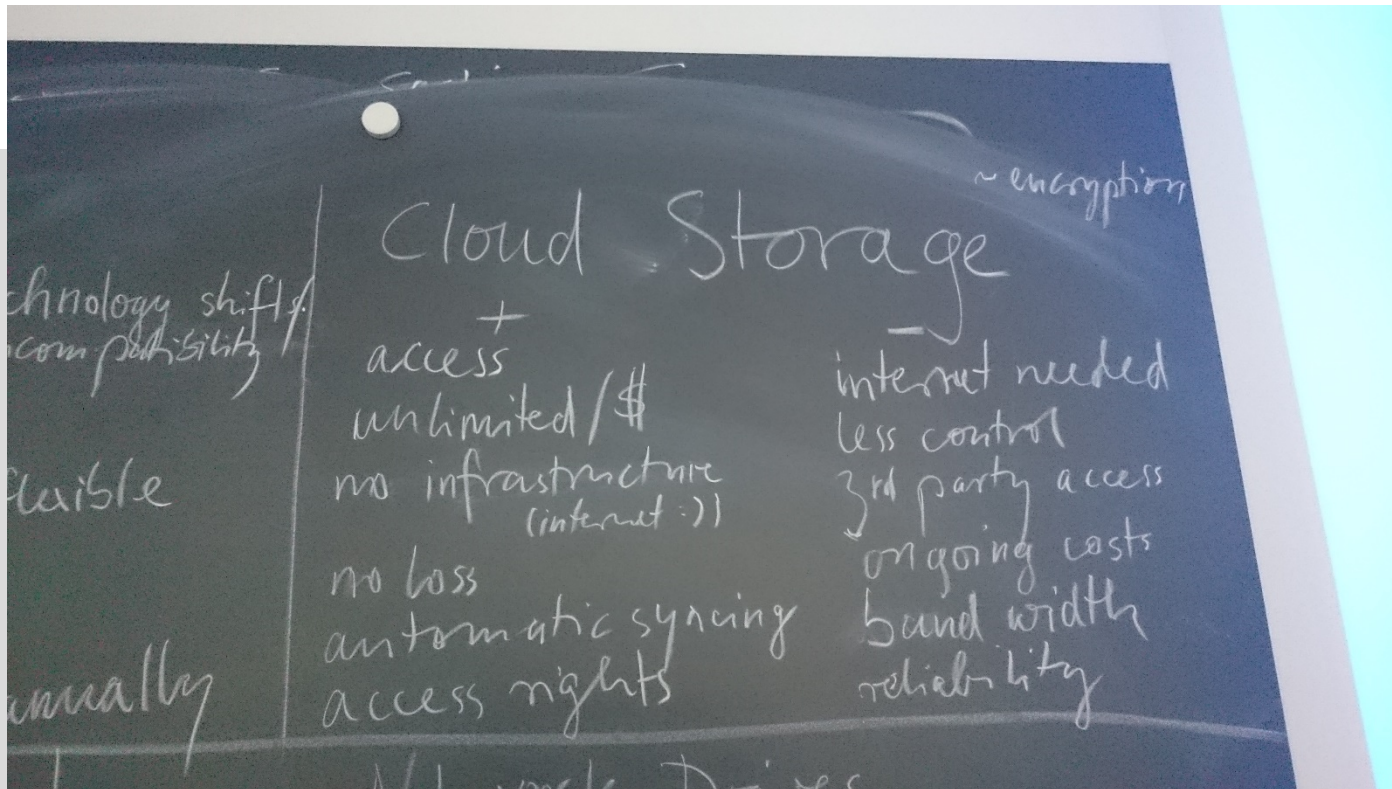
# Data storage

## Storage solutions – Local Storage

Local Storage		exchange +/- cost / manage manually	an a
+ encryption privacy availability fast, no internet just yours :)	- no sharing no portability limited space can be stolen location risks (fire, etc.) maintenance (e.g. backups) energy intensive noise compatibility	+ privacy collaboration expand / scale maintenance backup!	

# Data storage

## Storage solutions – Cloud Storage



# Data storage

## Storage solutions – Network Drives

manually	automatic syncing access rights	bandwidth reliability
Network Drives		
	+ privacy collaboration expand / scale maintenance backup	cost / investments same as local st. administrator reliability

# Data storage

## File formats and long-term storage

Not all formats are suited for archiving, if possible store data:

- In non-proprietary file formats
- Uncompressed
- Unencrypted

File formats for archiving:

- [ETH](#)
- [Kost](#) (german and french)



# Data storage

## Changing the file format

### Risks of file conversion:

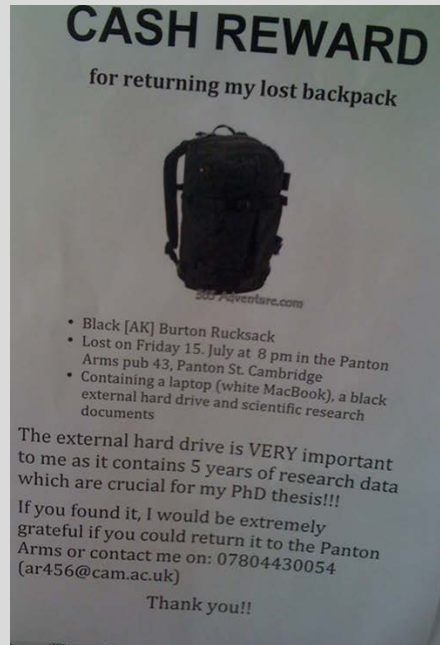
- loss of content
- loss of characteristics of the file stored within the file
- loss of layout
- loss of quality

- Not always possible
- Keep a copy in the original format
- Conversion recommendations: [ETH](#)
- Tools to validate the formats of your data files
  - [File Information Tool Set \(FITS\)](#)
  - [DROID](#)



# Backup

## Is it really necessary?



CC image by Sharyn Morrow  
on Flickr



CC image by momboleum  
on Flickr

# Backup

## Things to consider

- Who is responsible?
- What do you want to backup?
- How many backups and how frequently?
- Where will backups be stored?
- How much storage will you need?
- How will personal data be protected?
- Are there tools for automated backup?
- How long will backups be stored and how destroyed?
- How will you check the integrity of your backup files?

# Backup

## Simple rules

- Automated backups are better than manual
- 3-2-1 backup strategy: 3 copies, 2 different media, 1 external location
- Backups of sensitive data must be protected in the same way as the original files
- Regularly test whether restoring files from your backups is possible.
- Replace storage media regularly (portable storage media after 2-5 years)
- Tools for integrity checks: e.g. [MD5summer](#) or [Checksum Checker](#)



# Data Storage & Backup Media

## Optical

- » portable and low costs
- » small capacity, easily damaged and lost, not durable



## Portable Flash Drive

- » portable and low costs, robust and long-lived
- » small capacity and easily lost



## Some recommendations

- » use at least two types of storage media
- » replace storage media (after 2-5 years)
- » carry out integrity checks, e.g. by checksum tool

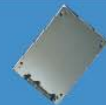
## Magnetic

- » low costs and high capacity
- » easily damaged, physical degradation



## Build In Flash Drive

- » robust and long-lived
- » high costs and small capacity



# Storage & Backup

## Take home

- You need a strategy
- Various storage solutions with advantages and disadvantages
- Not a one-size fits all solution: define a specific strategy for each project
- Short term strategy vs. Long term strategy

# Data sharing and reuse



# Data sharing and reuse

## Why?



<https://www.youtube.com/watch?v=jpGWfEgT0F0>, Odum Institute, CC BY

# Data sharing and reuse

Have you ever shared data?

Have you ever reused other people's data?

# Sharing Data

## Why?

- Funders' requirement
- Make use of resources more sustainable (reuse of data!)
- More transparency through reproducibility
- Access to data as a larger trend



# Sharing Data

## Why?

- Bring science forward, innovation
- Improvement and validation of research methods, better quality of data
- Increase impact and visibility of research
- Get more publications & citations
- Reuse other people's shared data



# Sharing Data

## How?

http://example.com

404

Page Not Found

<https://blog.algorithmia.com/404-error-scanner-algorithm-find-broken-links/>

The screenshot shows the top portion of a web browser displaying a Nature journal article. The browser's address bar shows a 404 error. The page header for Nature is visible, including navigation links and the article title. The article title is 'Publishers threaten to remove millions of papers from ResearchGate'. Below the title is a summary: 'Take-down notices “imminent” as lawsuit is filed alleging widespread copyright infringement.' The author is 'Richard Van Noorden'. The date is '10 October 2017 | Updated: 10 October 2017'. At the bottom of the article preview is a button labeled 'Rights & Permissions'.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

News & Comment | News | 2019 | May | Article

NATURE | NEWS

Publishers threaten to remove millions of papers from ResearchGate

Take-down notices “imminent” as lawsuit is filed alleging widespread copyright infringement.

Richard Van Noorden

10 October 2017 | Updated: 10 October 2017

Rights & Permissions

<https://dx.doi.org/10.1038/nature.2017.22793>



# Sharing Data

## How: some options

1. Data Paper / Data Journals
2. Data Repository

# Sharing Data

## Data Paper / Data Journal

- Useful for larger datasets
- Usually peer-reviewed
- In-depth description and contextualization
- one more publication
- Tip: link data paper to main paper record (DOI)
- Examples: [Foster](#), [HU Berlin](#)

Research Data Journal  
for the Humanities and Social Sciences

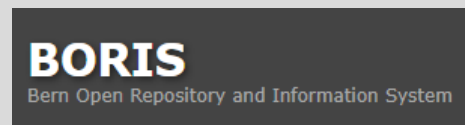
nature > scientific data

SCIENTIFIC DATA

# Sharing Data

## Repositories – general

- Online platforms
- Allow upload of files (e.g. research data)
- Describe and give proof of files
- Increase discoverability of deposits
- Generally: good repositories best way to share data

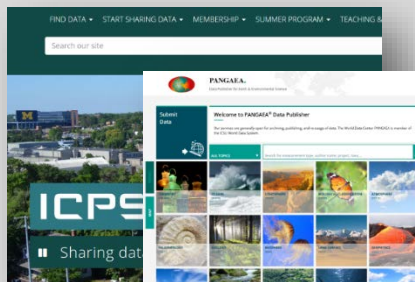


# Types of Repositories I

## Subject-specific

### Advantages

- Best visibility
- May offer subject-specific metadata
- Familiar with technical requirements for specific data



- **Examples:**
- [GenBank](#) (Genome data)
- [Pangaea](#) (Earth & Environmental Science)
- [ICPSR](#) (numeric social science data)

# Types of Repositories II

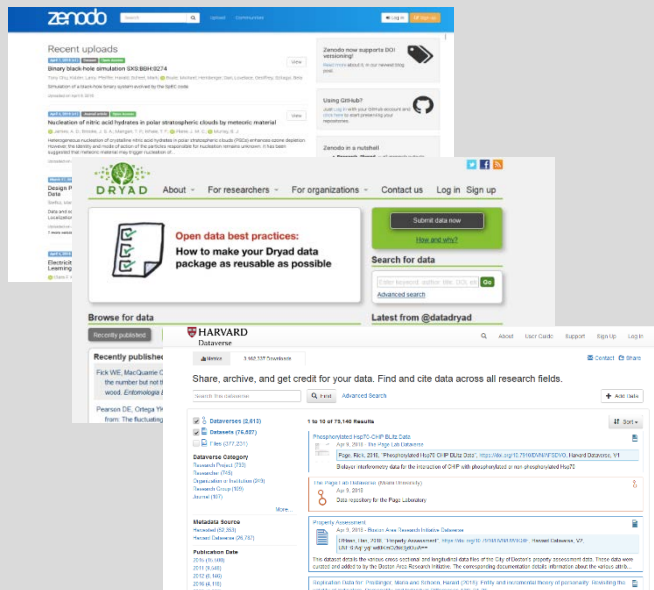
## Generic

### Advantages

- Open to all disciplines
- Easy to use

### Examples

- [Zenodo](#)
- [Harvard Dataverse](#)
- [Dryad](#)



# Types of Repositories III

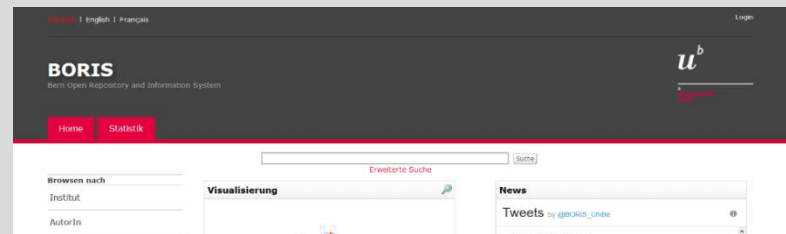
## Institutional

### Advantages

- Linked to your institution
- All Data at same location
- Financing of the repository is secured

### Examples

- BORIS for Publications  
([boris.unibe.ch](https://boris.unibe.ch))
- In process: BORIS Research Data



# How to find a suitable repository

Ask in your community

Search [re3data.org](https://www.re3data.org/)



# Data Reuse

## How?

- Find data via a search engine or on a repository
- Check thoroughly for copyright, licenses and other reuse regulations
- Restricted data (e.g. personal): data reuse agreements
- Act accordingly: mutual trust is the basis of data sharing and reuse!



# Data Reuse

## Data search engines

- [Google Dataset Search](#) (prototype)
- [Elsevier DataSearch](#) (prototype)
- [DataCite](#)

# Copyright & Licenses



# Copyright Transfer Agreement (CTA)

1.2 Author hereby grants and assigns to Publisher the exclusive, sole, permanent, world-wide, transferable, sub-licensable and unlimited right to reproduce, publish, distribute, transmit, make available or otherwise communicate to the public, translate, publicly perform, archive, store, lease or lend and sell the Work or parts thereof individually or together with other works in any language, in all revisions and versions (including soft cover, book club and collected editions, anthologies, advance printing, reprints or print to order, microfilm editions, audiograms and videograms), in all forms and media of expression including in electronic form (including offline and online use, push or pull technologies, use in databases and data networks (e.g. Internet) for display, print and storing on any and all stationary or portable end-user devices, e.g. text readers, audio, video or interactive devices, and for use in multimedia or interactive versions as well as for the display or transmission of the works or parts thereof in data networks or search engines, and posting the Work on social media accounts closely related to the Work), in whole, in part or in abridged form, in each case as now known or developed in the future, including the right to grant further time-limited or permanent rights. Publisher especially has the right to permit others to use individual illustrations, tables or text quotations and may use the Work for advertising purposes. For the purposes of use in electronic forms, Publisher may adjust the Work to the respective form of use and include links (e.g. frames or inline-links) or otherwise combine it with other works and/or remove links or combinations with other works provided in the Work. For the avoidance of doubt, all provisions of this contract apply regardless of whether the Work itself constitutes a database under applicable copyright laws or not.

# Copyright Transfer Agreement (CTA)

## 1.4

Author retains, in addition to uses permitted by law, the right to communicate the content of the Work to other scientists, to share the Work with them in manuscript form, to perform or present the Work or to use the content for non-commercial internal and educational purposes, provided the original source of publication is cited according to current citation standards.

# License

- Legal document that grants specific rights to users
- Copyright holder determines the conditions under which the work can be accessed, re-used and modified (if applicable)
- Removes any ambiguity over what others can – and can't – do with your data.
- Licenses can be applied to any material (e.g., articles, sound, text, image, multimedia, software) where exploitation or usage rights exist.
- License builds upon existing copyright regulations.

# Creative Commons















# Creative Commons

## Conditions




# Creative Commons Licenses


CREATIVE COMMONS LICENSES		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-NC	✓	✓	✗	✗	✓
	CC BY-NC-SA	✓	✓	✗	✓	✓
	CC BY-NC-ND	✓	✓	✗	✗	✓




You can redistribute  
(copy, publish, display,  
communicate, etc.)




You have to attribute  
the original work



You can use the work  
commercially



You can modify and  
adapt the original work



You can choose license  
type for your adaptations  
of the work.



# Creative Commons Licenses

## How does it work?



<https://vimeo.com/25684782>

# Exercise

What do you get?



**CC0** (Public Domain)

**CC BY** (Attribution)

**CC BY-SA** (Attribution + Share Alike)

**CC BY-ND** (Attribution + No  
Derivative)

**CC BY-NC** (Attribution + Non  
Commercial)

**CC BY-NC-SA** (Attribution + Non  
Commercial + Share Alike)

**CC BY-NC-ND** (Attribution + Non  
Commercial + Share Alike)

# Exercise

$$\text{CC BY} + \text{CC BY ND} = \text{CC BY ND}$$

$$\text{CC BY ND} + \text{CC BY NC} = \text{not allowed}$$

$$\text{CC BY} + \text{CC BY SA} = \text{not allowed}$$

# Creative Commons for Data



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

# How to License Research Data

DCC guide on [“How to License Research Data”](#)

[EUDAT Licensing Tool](#)

[Creative Commons License Chooser](#)

The screenshot displays the 'How to License Research Data' guide from the Digital Curation Centre (DCC) and JISC Legal. The interface is divided into several sections:

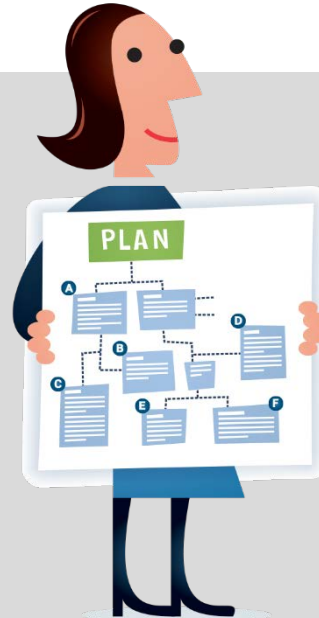
- Header:** A red banner at the top left reads 'A Digital Curation Centre and JISC Legal working level guide'. To the right are the DCC and JISC Legal logos.
- Main Title:** 'How to License Research Data' in large red font, with 'Alex Ball (DCC)' listed below it.
- License Features:** A section titled 'License Features' with a sub-header 'Your choices on this panel will update the other panels on this page'. It contains two questions with radio button options:
  - 'Allow adaptations of your work to be shared?' with options: Yes, No, Yes, as long as others share alike.
  - 'Allow commercial uses of your work?' with options: Yes, No.
- Selected License:** A section titled 'Selected License' showing 'Attribution 4.0 International' with the CC BY icon. A note below states 'This is a Free Culture License!'.
- Interactive Questions:** Two pop-up boxes with 'Yes' and 'No' buttons:
  - 'Do you own copyright and similar rights in your dataset and all its constitutive parts?'
  - 'Do you allow others to make commercial use of you data?'
- Footer:** Includes the Creative Commons Attribution (CC-BY) license description and the Public Domain Dedication (CC Zero) license description.

# Wrap up

## Data Management Planning

*u<sup>b</sup>*

<sup>b</sup>  
UNIVERSITÄT  
BERN



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

# Write a DMP









## Common themes in a DMP

- Description of data to be collected / created
- Standards / methodologies for data collection
- Data organization and file naming
- Ethics and Intellectual Property
- Short- and long-term storage and backup
- Data sharing

# Need any help?

## Check out our website!

[www.unibe.ch/ub/openscience](http://www.unibe.ch/ub/openscience)

<p>OPEN ACCESS</p>  <p><b>OA</b></p> <p>Here you find an overview of the subject area Open Access.</p>	<p>SERVICES</p>  <p><b>Services</b></p> <p>Learn about our many services: information, training and support.</p>	<p>DISSERTATIONS</p>  <p><b>Publish online</b></p> <p>Learn how to publish your doctoral thesis online and open access.</p>	<p>BORIS</p>  <p><b>BORIS Publications</b></p> <p>Here you find information about the institutional repository of the University of Bern.</p>
<p>BERN OPEN PUBLISHING</p>  <p><b>Journals &amp; Series</b></p> <p>Here you find technical and administrative support for publishing books and journals.</p>	<p>RESEARCH DATA MANAGEMENT</p>  <p><b>Data Management</b></p> <p>Here you find information about research data management and DMPs.</p>	<p>LONG TERM PRESERVATION</p>  <p>How can data produced at the University of Bern be digitally archived?</p>	<p>IDENTIFIERS</p>  <p><b>ORCID &amp; Co</b></p> <p>ORCID iDs, DOIs, ISBNs and ISSN make you and your research unique.</p>