Foreword

The workshop will begin at 09:00

The slides from this workshop can be dowloaded at :
https://www.unige.ch/researchdata/en/support/all/index/

DOWNLOADS

Download the slides of the workshop **How to fill the DMP of the SNSF?**

Download the slides from the lectures of the 4th Sept. 2018 during the **Talk with the SNSF**

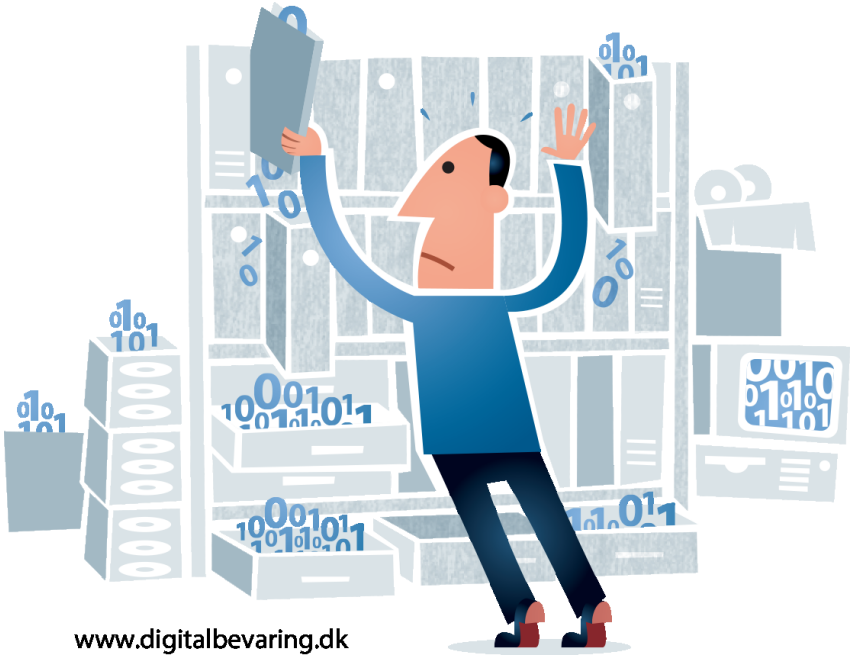
Download the slides of the *Midi de l'info* **Sharing your data?** (spring 2019)

Access the slides of the SOPSI's Open Science Day, (July 1st, 2019)

Download the slides of the **Managing research data as a junior scientist** workshop from *the transversal PhD Library Camp of Graduate Campus* (September 2020)

Section «Download», (at the very bottom of the page)

SNSF Data Management Plan (DMP)



How to fill it and what information to include ?

Lydie Echernier, Claire Wuillemin

September 2020

Division de l'information scientifique

Université de Genève

Plan

1. General principles about Research Data and SNF requirements
2. What is the Data Management Plan (DMP)?
3. The essentials in a nutshell
4. Our services to create your DMP

UNIGE's policy

- ✓ Approved by the Rectorat after consultation with the faculties and deans (June 2018)
- ✓ 1st university in Switzerland with a policy on the management of research data
- ✓ Establishment of responsibilities :
  - of the University
  - of researchers

https://www.unige.ch/researchdata/en/services/all/policy/

1

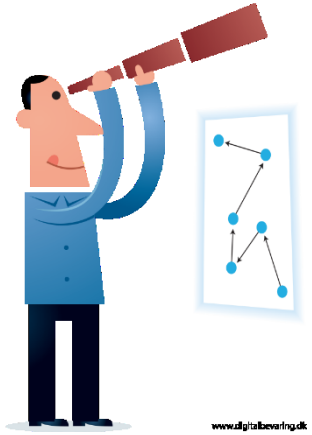
What are research data?

Factual records (numerical scores, textual records, images and sounds) used as **primary sources** for scientific research, and that are commonly accepted in the scientific community as **necessary to validate research findings**

OECD's Principles and Guidelines for Access to Research Data form Public Funding, 2007, p.18.
http://www.oecd.org/fr/sti/sci-tech/38500823.pdf

Classification of research data

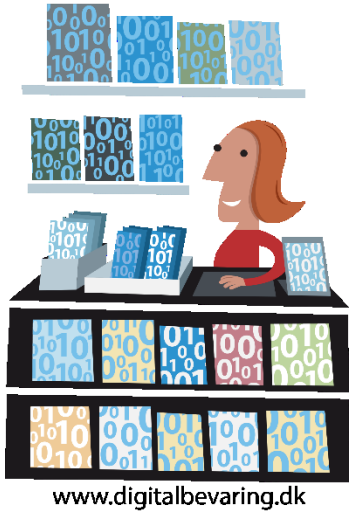
observational



simulation / models

derived / compiled

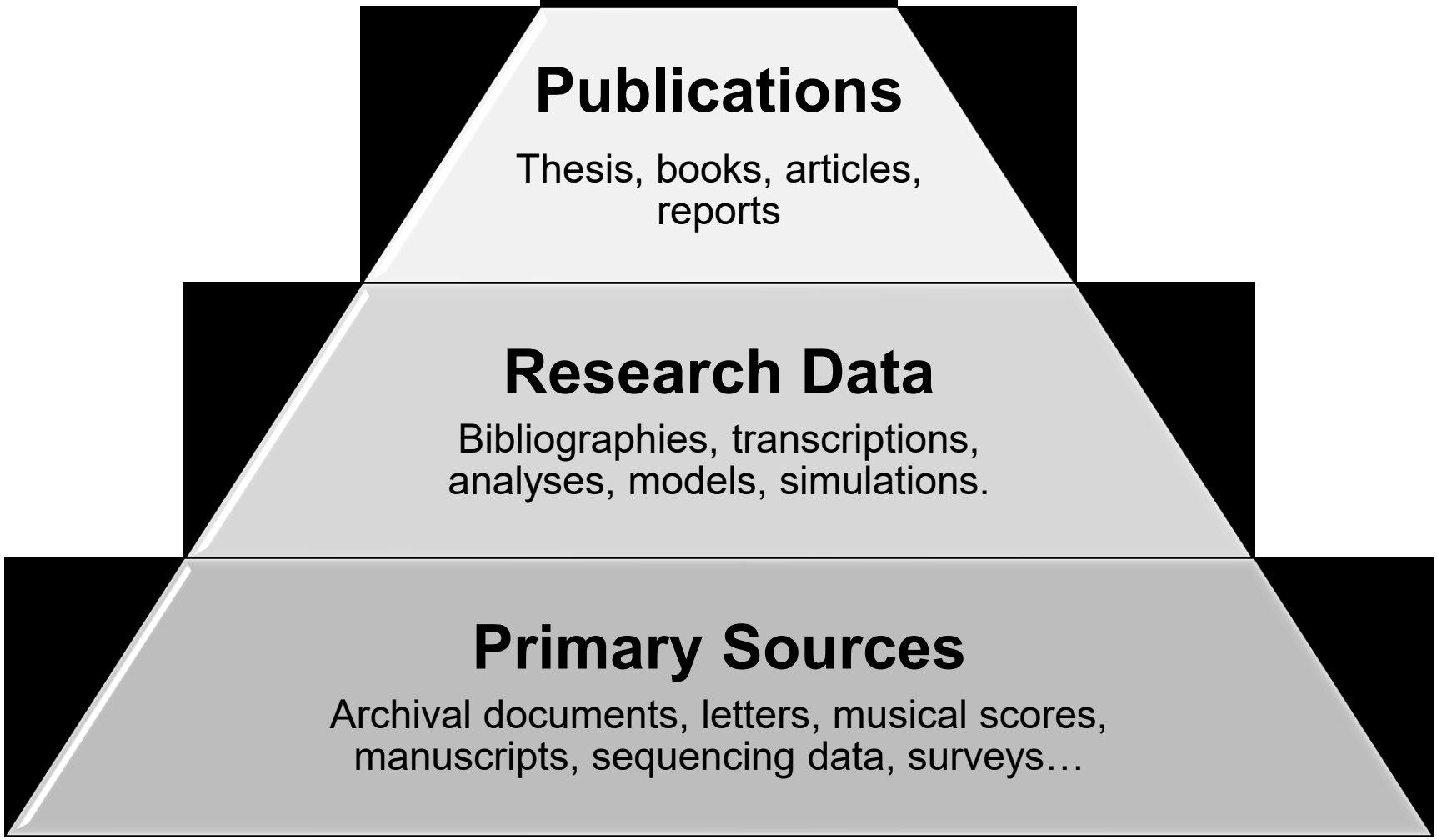
experimental



reference

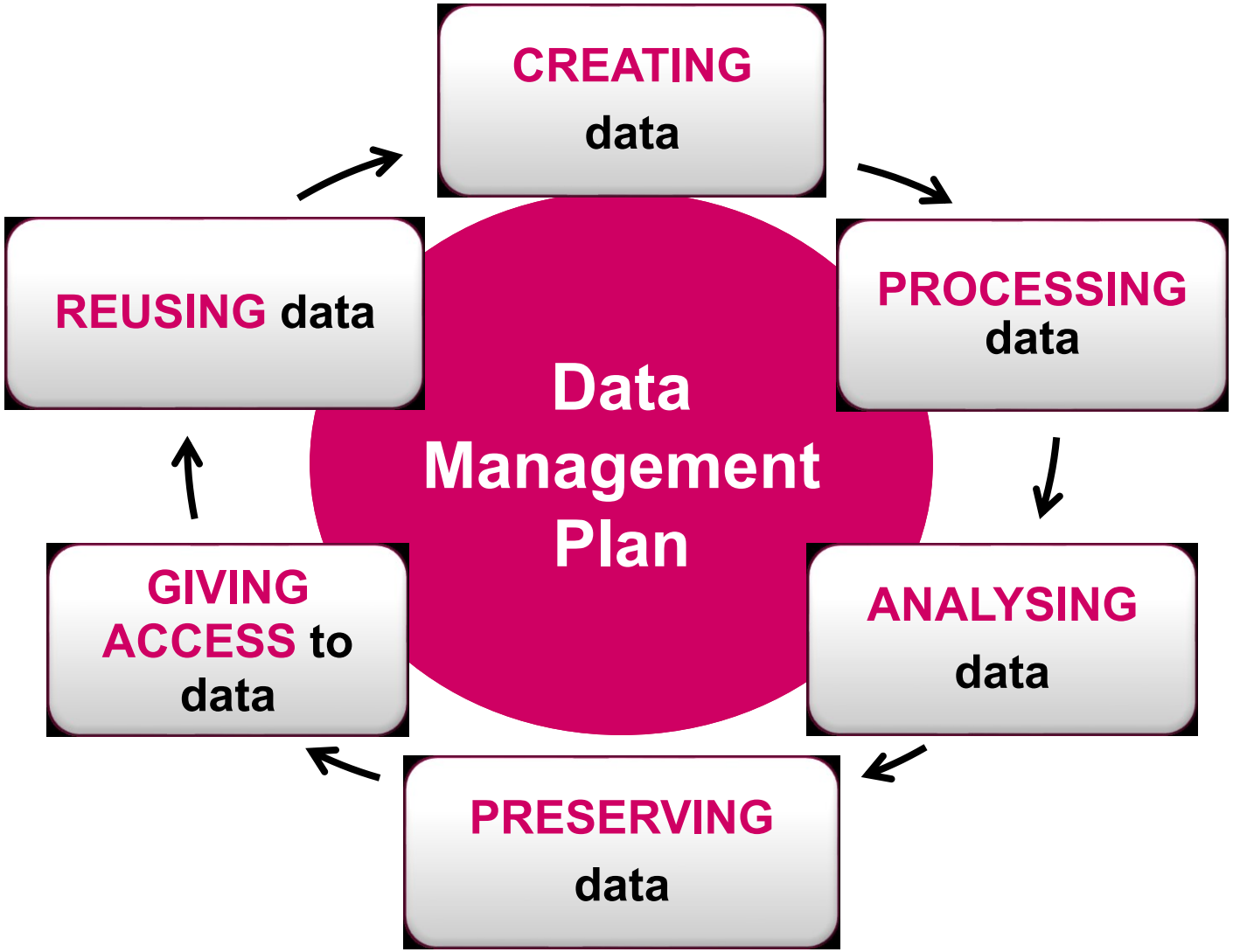
1

Sources → Data → Publications

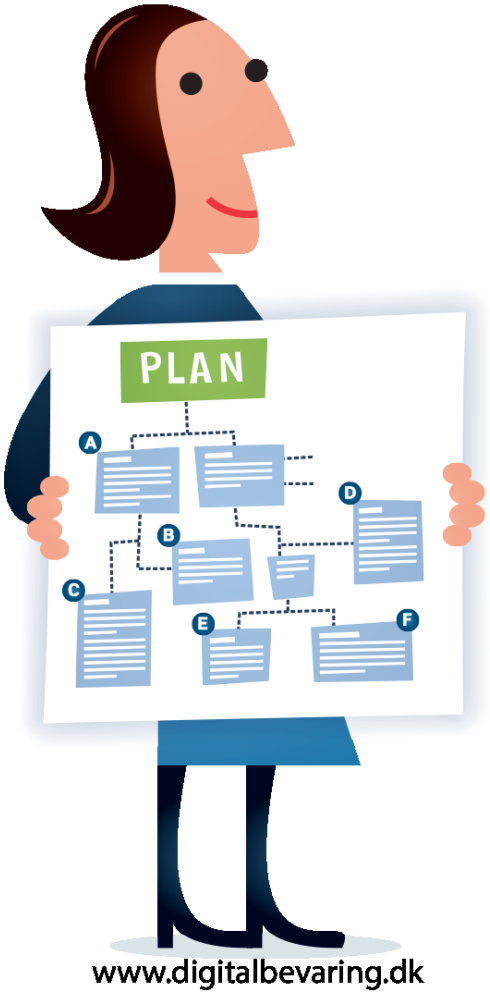


Selon : Andorfer P. *Forschen und Forschungsdaten in den Geisteswissenschaften: Zwischenbericht einer Interviewreihe*: Niedersächsische Staats-und Universitätsbibliothek Göttingen; 2015.

Data lifecycle



Data Management Plan



- Saves **time**
- Increases the **impact** of your research
- Participates in **open science**
- Facilitates new **discoveries**

2

Requirements from SNF

mySNF

2. Application Data

#	No./Title	Status
2.1	Basic data I	In preparation
2.2	Basic data II	In preparation
2.3	Use-inspired project	In preparation
2.4	Re-submission	
2.5	Continuation of	
2.6	Link to other SNSF projects	
2.7	Further requested and available funds (not fron	
2.8	University or research institution	
2.9	Requested funding	
2.10	Data management plan (DMP)	
2.11	Research requiring authorisation or notification	
2.12	Exclusion of external reviewers	
2.13	General remarks on the project	

.. Données relatives à la requête

#	No/titre
2.1	Données de base I
2.2	Données de base II
2.3	Projet orienté vers l'application
2.4	Resoumission
2.5	Continuation
2.6	Relation avec d'autres projets FNS
2.7	Autres moyens demandés et disponibles (qui ne sont pas du FNS)
2.8	Haute école
2.9	Besoin financier
2.10	Plan de gestion des données (DMP)
2.11	Recherche exigeant des autorisations ou des annonces

2

The DMP on MySNF

I do not submit a DMP for the following reason:

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.

1. Data collection and Documentation

☐ 1.1 What data will you collect, observe, generate or reuse?

List of data :

→ *Type*

→ *Format*

→ *Volume*

2

1.1 What data will you collect, observe, generate or reuse?

Existing data to be reused in the project consists of **transcribed texts**, **facsimile images**, and **audio files** found in the third-party database Australian Drama Archive

https://www.austlit.edu.au/australiandramaarchive

The project itself will generate the following types of data:

- **Text transcriptions** from extracts of earlier plays which do not already exist in the Australian Drama Archive. These will be produced in the **Open XML format .docx** during the in-project period.
- **Images**. Where relevant the project will produce **photographic images** of some of the features of the original playscripts or ephemera identified during archival trips. An **estimated 500 images** will be taken in high resolution (2000 X 3000 pixels) as **.jpg files**.
- **Tables and figures** in **.xlsx format**, outlining different characteristics of monologues found in those earlier plays.

It is expected that the total volume of data generated **will not exceed 20 GB** but this will be further defined once archival trips are completed at the end of the first year of the project.

2

1.1 What data will you collect, observe, generate or reuse?

Qualitative (e.g. phenotyping data for cell lines), **quantitative** (e.g. cell counts), mass spectrometry and **image** data will be generated. Raw data will be analysed and expressed as graphs, tables and annotated images.

Data generated will be in **various formats and sizes of datasets**. They include:

1. **Cell images** e.g. phase and fluorescence, and electron micrographs (**~5,000 images** over project). Software used includes OpenLab, Softworx and IN Cell Investigator, with data saved as **software-specific files e.g liff and lg3 files**, as well as generic formats such as **jpeg, tiff** etc.
2. **mass spectrometry spectra** (from **<50 samples**). MS data will be analysed using Bruker Data Analysis or Thermo Excalibur software (generating **xml and raw files**) and proteins will be matched to the *T. brucei* genome dataset using the Matrix Science Mascot search engine. Each LC-MS data file is between **1-2GB.**
3. Cell line phenotyping data including growth curves and DAPI counts (**Excel and GraphPadPrism files**) and flow cytometry data (FlowJo and **jpeg/tiff files**) (~200 data sets).
4. …

(source)

1. Data collection and Documentation

☐ 1.2 How will the data be collected, observed or generated?

Data processing :

→ *Standards*

→ *Methods*

→ *Files naming and versioning*

2

1.2 How will the data be collected, observed or generated?

All samples on which data are collected will be prepared according to **published standard protocols in the field**.
Quality of analytical data will be guaranteed through **calibration of devices** and comparison with internal standards. Appropriate experimental design, data recording and data validation (controls, randomization/ blinding, sampling/replicates, experimental versus hypothesis driven-protocol) will be used, ensuring internal validity.
Methods and materials will be recorded using an **Electronic Lab Notebook (ELN).** All experimental data will be automatically imported into a **Laboratory Information Management System (LIMS)** from the measurement device.

Files will be named according to a pre-agreed **convention** including a short, unique identifier for the project, a summary of the content, the date (YYYYMMDD format) and the document version. Ex. : **YYYYMMDD_ProjectIDX_ContentSummary_V01**

Peer review will be employed as an additional step of quality control in the form of **regular supervision and lab meetings**.

(eg from DMP Canva Generator, VitalIT)

1. Data collection and Documentation

1.3 What documentation and metadata will you provide with the data?

→ *Information required for future users*

→ *Annotations*

2

1.3 What documentation and metadata will you provide with the data?

A metadata file (named README) will be saved in the root folder of the project. It will contain the following information:

- The **names** of the Principal investigator and co-investigators
- A description of **methodology and tools** used for data collection
- The **dates** of data collection
- A description of the **directory hierarchy**, of the **type of data** it contains and the **filenaming convention**
- A **complete list** of any headings/codes/abbreviations and conventions used in the files
- The **sharing and accessing conditions** (license) during and after the end of the project

Metadata sources

- Use existing metadata standards whenever possible
- Document your metadata scheme as soon as possible
- Examples: author, title, date, brand and model of the experimental apparatus…

http://www.dcc.ac.uk/resources/metadata-

https://fairsharing.org/

- How to write a good README.txt

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

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

→ *Data protection*

→ *Confidentiality agreement*

→ *Personal and sensitive data management*

2

2.1 How will ethical issues be addressed and handled?

All interviewees and focus group participants will sign a **Consent form** agreed to by **UNIGE ethics committee**.

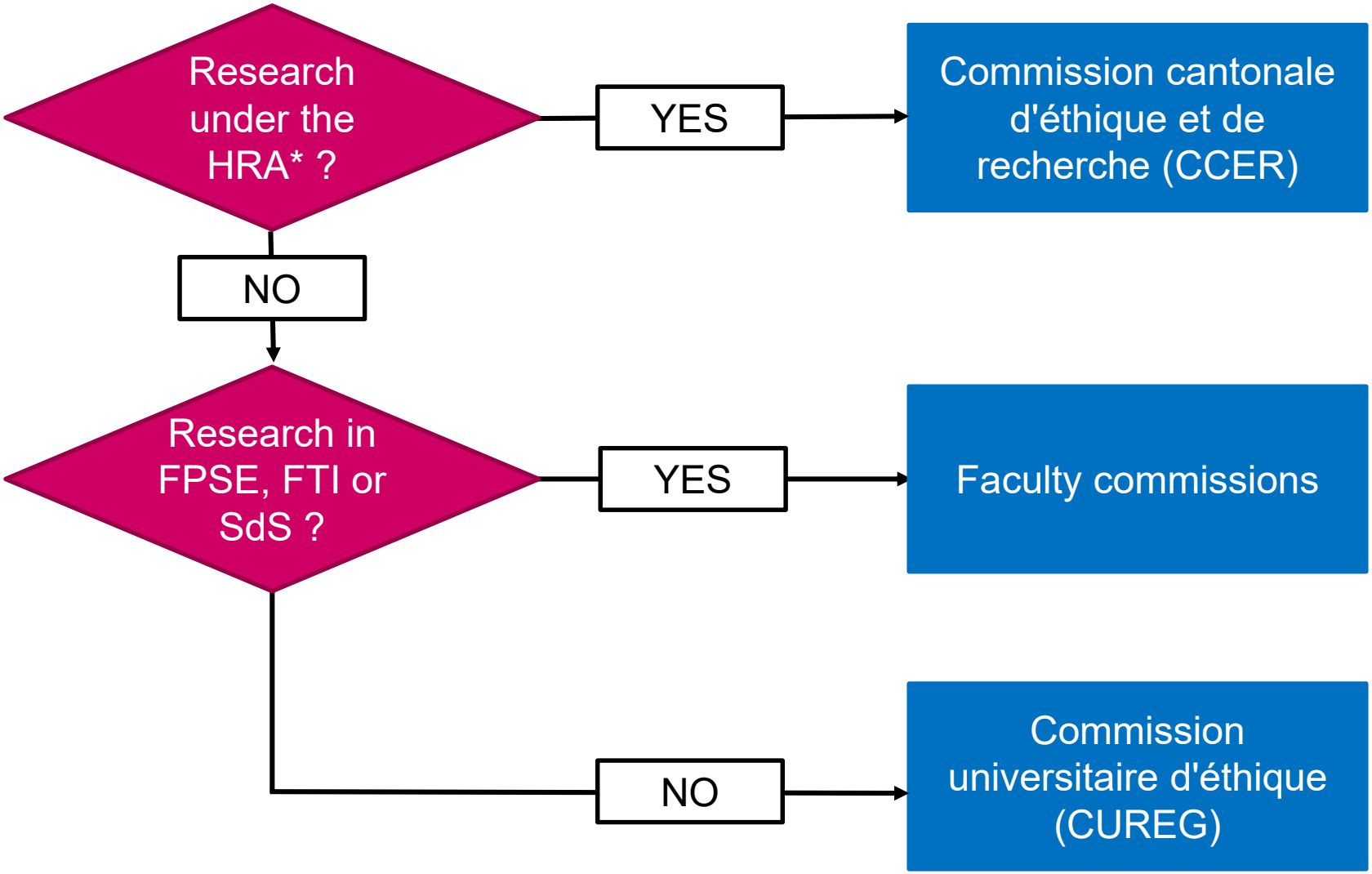
We have guaranteed anonymity to our interviewees and focus group participants. Therefore we will not be depositing .wav files as this would compromise that guarantee.

However, **anonymised transcripts** of the interviews and focus groups will be deposited. Unless such anonymisation is impossible or associated with a disproportional effort, in which case part or all of the data will not be shared publicly. We will make sure consent forms make provision for future sharing of data.

(source of this eg., adapted)

2

Which ethics committee?



* 810.30 - Swiss Federal Act on Research involving Human Beings

2

UNIGE's guidelines

- ✓ **Code of ethics and professional conduct**

  "*Any individual engaged in research must ensure that integrity is maintained in his or her quest for knowledge, in the interpretation and application of the results and in the structuring of his or her research*." (article 1, letter c)

- ✓ **Research integrity directive**

  The head of the research project must ensure that the essential data are kept securely for at least five years after the end of the project. (point 2.6)

- ✓ **Swiss Federal Act on Research involving Human Beings (HRA)**

- ✓ **Commission cantonale d'éthique de recherche (CCER)**

- ✓ **Commission universitaire d'éthique (CUREG)**

- ✓ **Faculty Committees**
  - Psychologie et Sciences de l'éducation
  - Traduction et Interprétation
  - Sciences de la Société

2. Ethics, legal and security issues

2.2 How will data access and security be managed?

Risks management:

→ *Data security*

→ *Digital*

→ *Material*

→ *Data access rights and permissions*

2

2.2 How will data access and security be managed?

The **main risks** to data security are loss or damage to laboratory notebooks and loss or corruption of electronic data. Data will be safeguarded by the following measures:

1. Data in **lab notebooks** will also be recorded in electronic form that is backed up daily to secure against loss or damage of the notebook.
2. Access to electronic data (prior to publication) will be limited to the members of the research group and relevant collaborators via **limiting access to shared drives on the University server**.
3. Access to laboratories and offices are controlled by **card access** to reduce the likelihood of malicious loss/damage; all computers used in this project will run Standard Staff Desktop, whereby **firewalls and antivirus software** are automatically upgraded; staff will **lock their workstation** whenever they are away from it.

(source of this eg., adapted)

2. Ethics, legal and security issues

☐ 2.3 How will you handle copyright and Intellectual Property Rights issues?

Legal aspects:

→ *Who is the owner of the data?*

→ *Which licence to apply?*

2.3 How will you handle copyright and Intellectual Property Rights issues?

Some specific published editions of texts analysed in this project are **proprietary**, and not available for release.

In those cases **specific agreements will be made with the collection owners** allowing public access to a limited number of images and text extracts through the web site and other interfaces described in the project proposal.

Advice from the University Legal Office will be sought as necessary.

New **data derived from analysis** of those texts carried out during the project (tables, figures, annotations...) will be made **publicly accessible under a CC-BY license**.

2

Choose your license

Data under license/contract?

- yes → Respect license/contract Contact Unitec for help
- no ↓

Contains personal/sensitive data?

- yes → Check anonymisation before sharing!
- no ↓

Potential commercial use?

- yes → Contact Unitec for help
- no ↓

Allow fully open reuse?

- no → Contact Library
- yes ↓

Publish data under **CC0** or **CC-BY** license

See also: https://creativecommons.org/choose/

3. Data storage and preservation

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

During the project:

→ Storage capacity

→ Storage facility

→ Back-up procedures

2

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

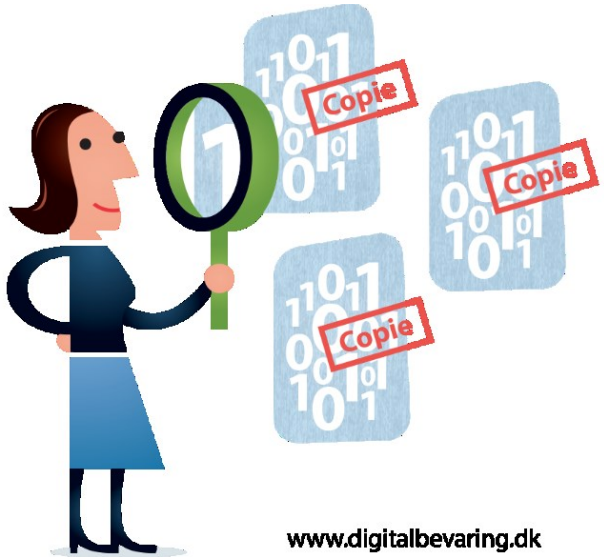
During the project, data will be stored on the **PI's computer hard drive** and backed up weekly:

- On an **external hard drive** kept in a locked filing cabinet in the PI's office – the PI will perform the backup
- On the **academic NAS managed by the University of Geneva's IT department**. This academic NAS follows common protocols and best practices to ensure maximum security, integrity and availability. It extends over two distinct physical locations (UniDufour and Campus Biotech) and automatically performs a snapshot of files every 4 hours, with a retention of copies of 6 weeks.

Consent forms and all identifying information will be kept in a locked filing cabinet and not stored with electronic files.

Good practices for data storage

→ Store **3 copies** of your data



1. The original
2. A copy kept on a local external device
3. A copy kept on an external device at a different location

UNIGE's NAS servers, a shared storage space, covers points 2. and 3.

3. Data storage and preservation

3.2 What is your data preservation plan?

After the project's end:

→ Procedures to select data to be preserved

→ File formats for preservation

2

3.2 What is your data preservation plan?

Audio recordings of interviews and **transcripts** will be kept in .wav and .pdf format for a minimum of 10 years on a hard drive in a locked filing cabinet in the PI office.

Consent forms will be kept for 10 years in the same locked filing cabinet.

However, **anonymised transcripts** of the interviews and focus groups will be distributed under a CC-BY license **in .csv format**.

Data that cannot be anonymized will not be shared publicly.

4. Data sharing and reuse

4.1 How and where will the data be shared?

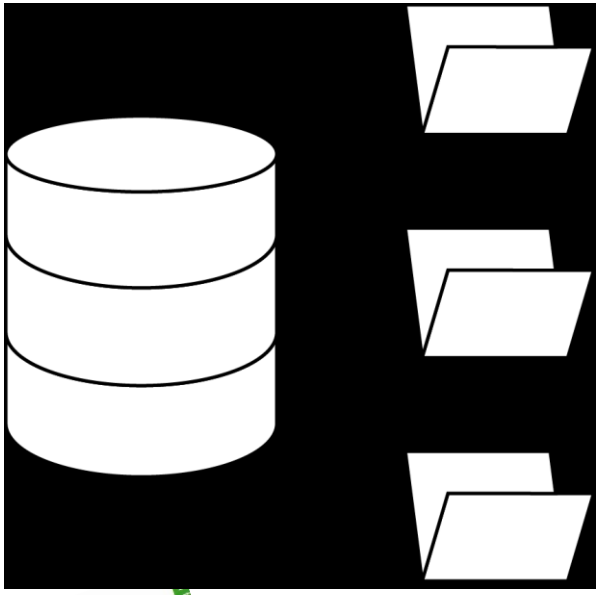
Archival and publishing:

→ Which repository to select

→ How others will find out about your data?

Archiving and dissemination

Repository

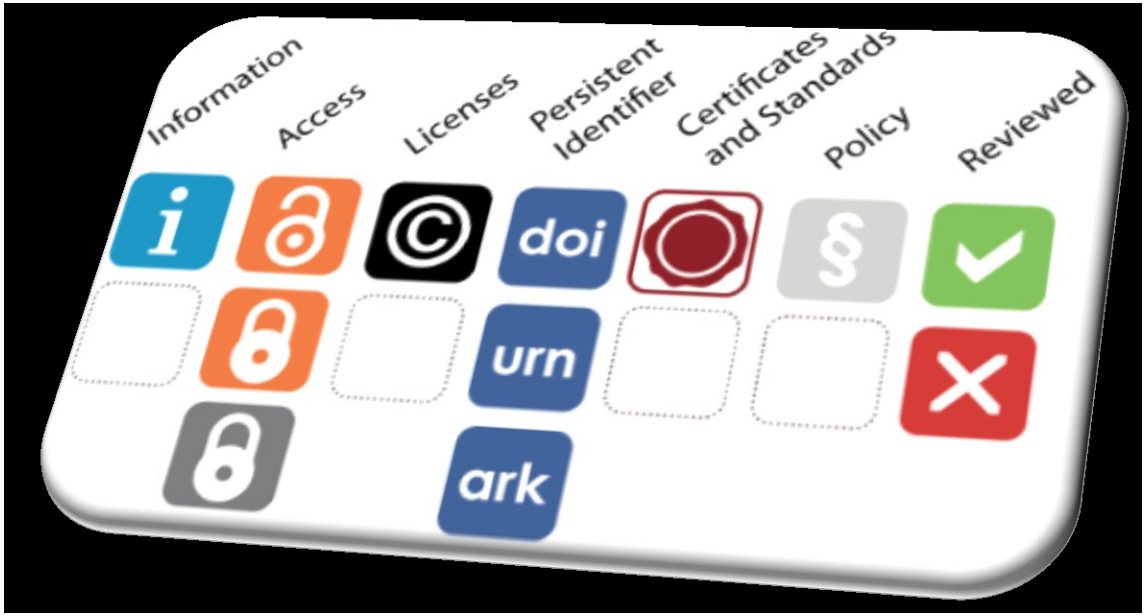


- ✓ general purpose (multi-disciplinary)
- ✓ discipline-specific
- ✓ institutional

- Journal supplementary material service
- Departmental, project or personal web page

Trouver un repository

- **Yareta**
- Zenodo
- OSF.io
- Dryad
- …



For other repositories, check: re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

Yareta

- ✓ Developed as part of a cantonal bill
- ✓ Repository for the researchers of Geneva
- ✓ Swiss-based servers (original et copy)
- ✓ Compliant with **FAIR principles &** provides **DOI**
- ✓ **Free** up to 50 Go
  (above: 100 CHF / Tbit / year of retention)

https://yareta.unige.ch

2

4.1 How and where will the data be shared?

The project data will be shared with **Yareta**, the research data repository of Geneva's Higher Education Institutions, Along with the README file mentioned in 1.3.

The DOI issued to datasets in the repository will be included as part of a data citation in publications, allowing the datasets underpinning the publication to be easily identified and accessed.

4. Data sharing and reuse

☐ 4.2 Are there any necessary limitations to protect sensitive data?

→ Under which conditions will the data be made available?

→ Timing of data release

→ Delay if applicable

2

4.2 Are there any necessary limitations to protect sensitive data?

Individual research subjects' data cannot legally nor ethically be made available to non authorised people (HRA, *cf.* §**2.1**). Only the sponsor, the investigation team, reviewers, auditors and inspection authorities are entitled to access such data.

No personal data or data that may easily identify subjects will be provided, with respect to the Swiss law on human research (Federal Act on Research involving Human Beings (HRA)) and its applicable ordinance ClinO/KlinV/OClin/OSRUm.

(eg from DMP Model, HUG CRC)

4. Data sharing and reuse

☐ 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.

→ ☐ *[checkbox]*

☐ 4.4 I will choose digital repositories maintained by a non-profit organisation.

→ *yes* / no *[radio button]*

2

Is my repository FAIR ?

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Findable **A**ccessible **I**nteroperable **R**eusable

At least **one blue icon**

At least **one orange locker**

Better follow the standards

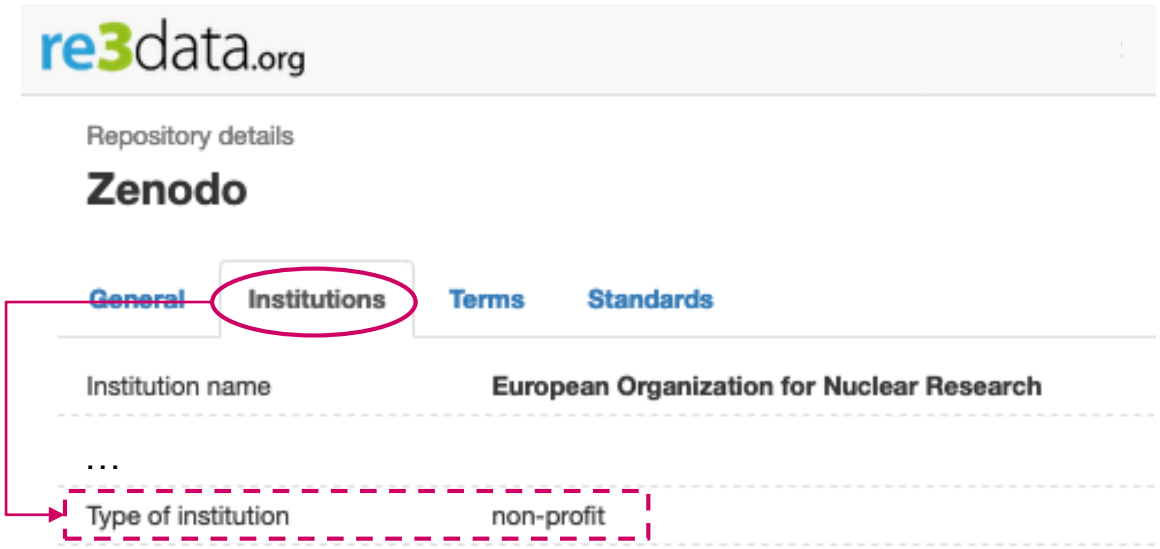
Licence icon **mandatory**

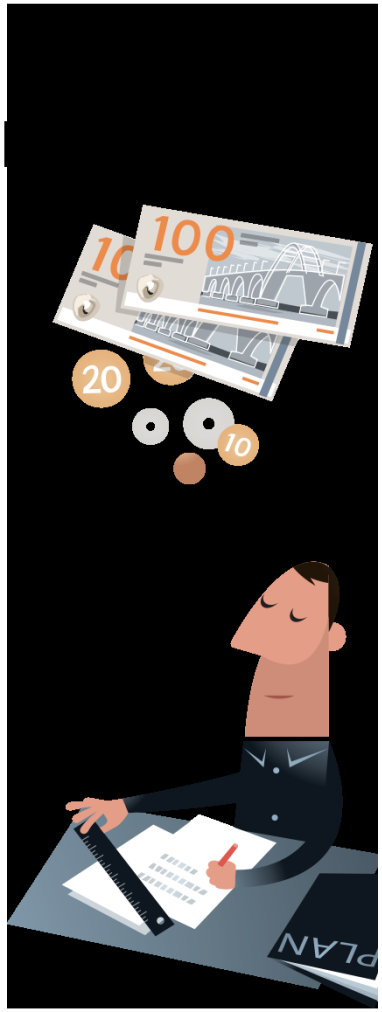
2

Data management costs

The SNF allows you to ask for up to **10’000.- CHF** for:

- Data preparation and validation (**stewards**
- **Uploading** your data on a non-commercial FAIR repository



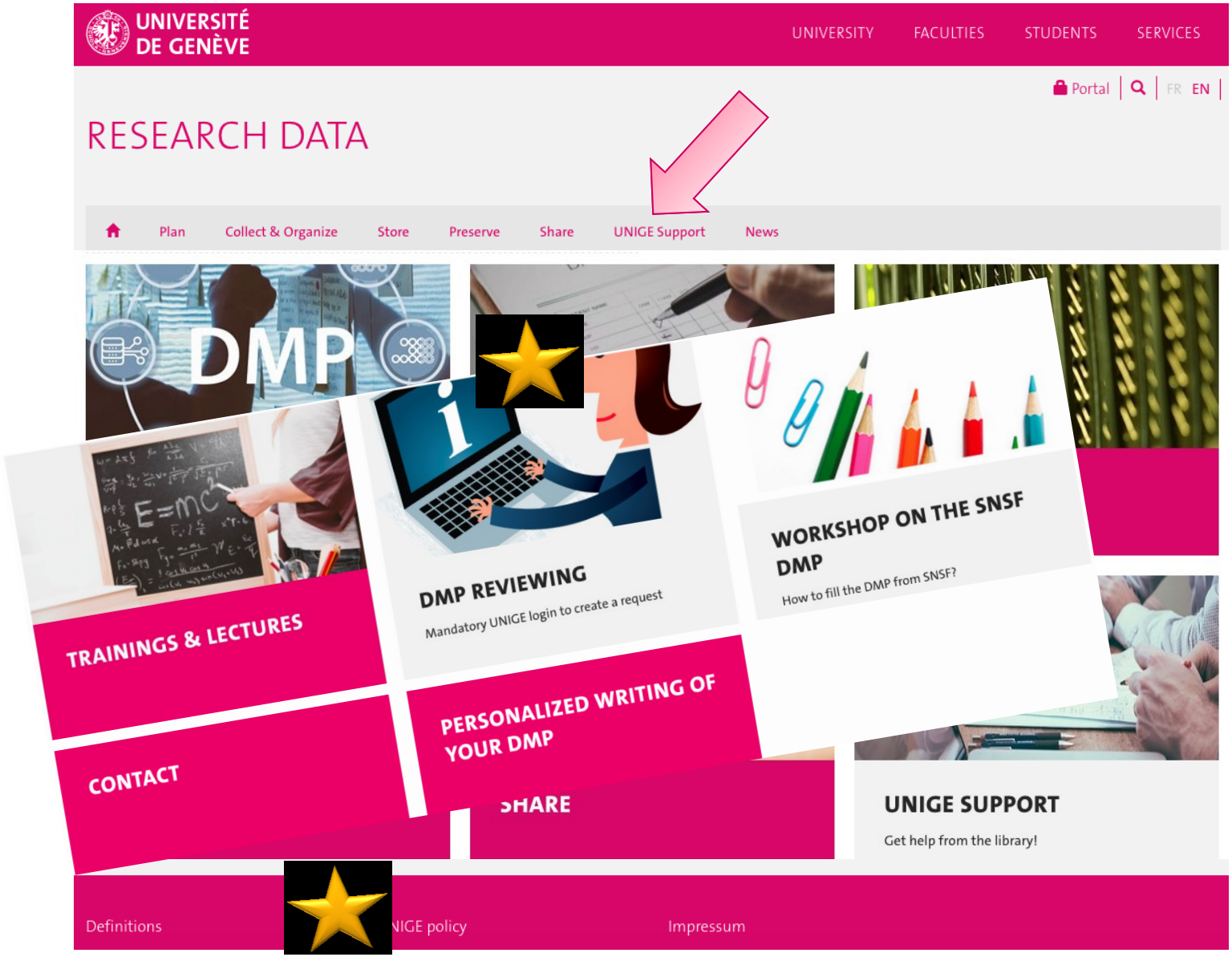


Take home message

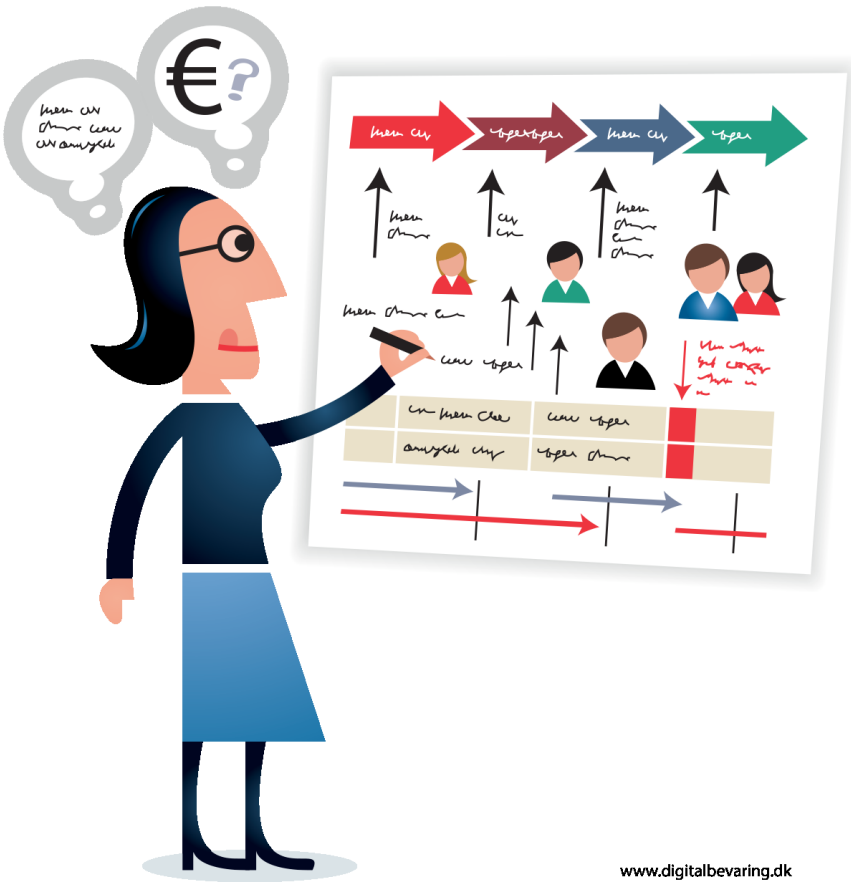
- ✓ A “plausible” DMP is a *sine qua non* condition for the release of the funds
- ✓ DMPs are editable throughout the funding period (and must be updated)
- ✓ Once SNSF funding has ended and the final scientific report has been approved, the DMP cannot be modified anymore
- ✓ The DMP is shared on P3 (SNSF’s public database) at the end of the project

4

www.unige.ch/researchdata



Drafting your DMP



Date : Friday, 25 September

Hours : from 08h30 to 12h00

Venue : Uni Mail, room M2220

Language : English and French

During this workshop, there will be no theoretical presentation, but only individual work (yours) and assistance provided by specialists.

Times are given as an indication:
you may join us as briefly as you wish, based on your needs.

Online registration

Contacts

For more informations or help on DMPs and research data:

researchdata-info@unige.ch

http://www.unige.ch/researchdata

Illustrations : CC BY Jørgen Stamp, http://digitalbevaring.dk