



Gestion des données

Ça commence avant d'en avoir

Guilaine
Baud-Vittoz
&
Francesco Varrato

HES-SO

09.05.2022

- A Marie-Laure Berchel et la HES-SO de nous avoir invité
- A tous nos collègues qui permettent le soutien aux données de la recherche possible jour après-jour

DATA ou Données

Factual records: numerical scores, textual records, images, sounds, protocols, source code, ...

RESEARCH DATA (RD) ou Données de recherche (DR)

Data used as primary sources for scientific research, and commonly accepted in the scientific community to validate research findings (OECD)

RESEARCH DATA MANAGEMENT (RDM) ou Gestion des Données de la recherche (RDM)

The care and maintenance of research data during the research cycle (UC Berkeley Library)

OPEN RESEARCH DATA (ORD)

Research data that managed during its **lifecycle** to comply with the **FAIR** principles



RDM: Research Data Management

DMP: Data Management Plan

ORD: Open Research Data

OS: Open Science

SNSF: Swiss National Science Foundation

ERC: European Research Council

GDPR: General Data Protection Regulation

ELN: Electronic Laboratory Notebook

etc...



FINDABLE = **T**ROUVABLE / **R**ECHERCHABLE

Data and metadata are easy to find by both humans & computers

ACCESSIBLE = **A**CCESSIBLE

Machines & humans can readily access or download (meta)data

INTEROPERABLE = **I**NTEROPÉRABLE

Data from different datasets are ready to be read or combined

REUSABLE = **R**ÉUTILISABLES

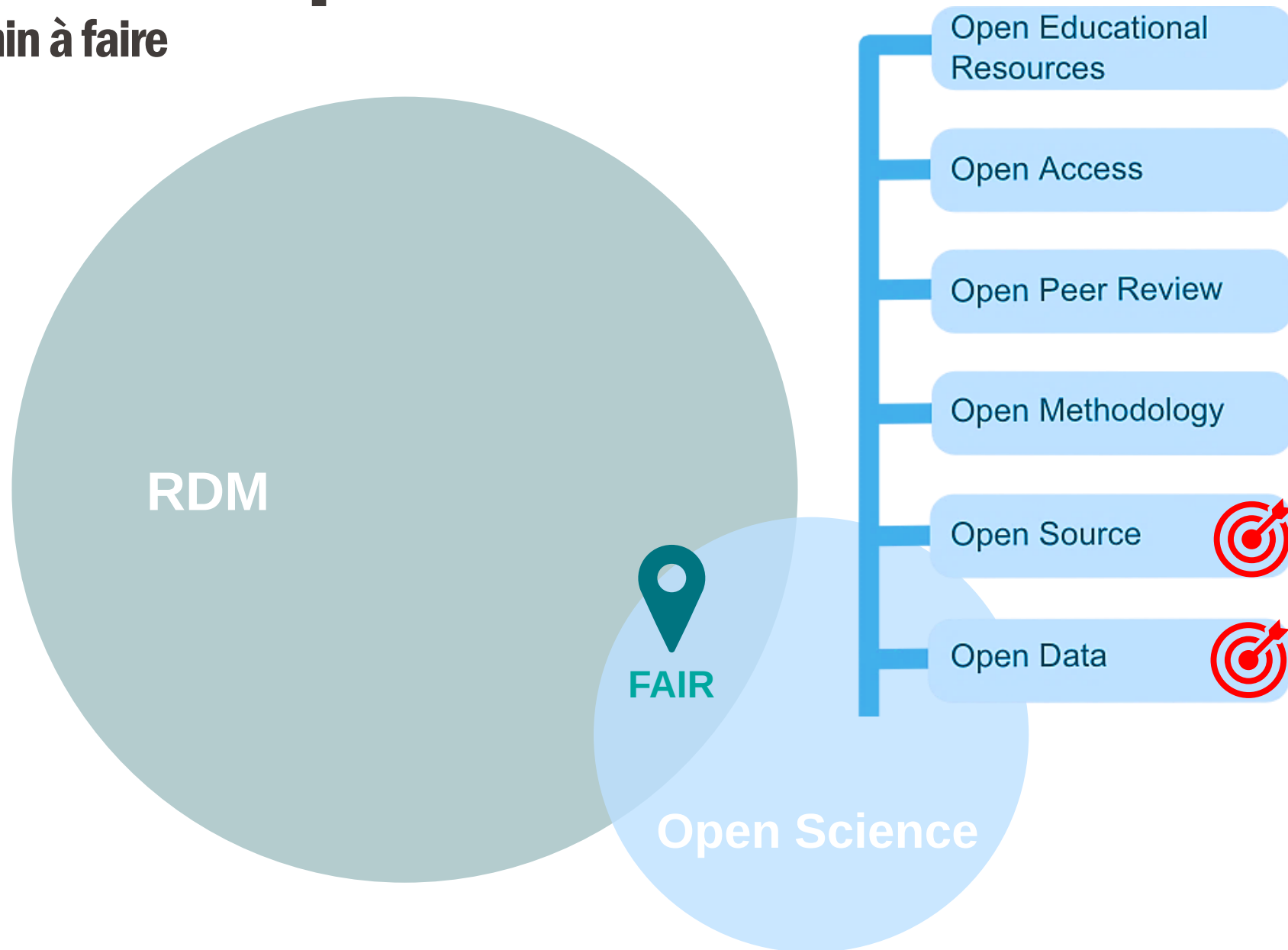
Data and metadata are easily replicated or combined in future research

Check out:

- [Scholarly publications and the FAIR Principles](#)
- [Turning FAIR into reality](#)
- EPFL [FAIR Data Principles FastGuide](#)
- [How to make your data FAIR](#)

RDM → FAIR! → Open Data

encore du chemin à faire



Open Data → FAIR → RDM

RDM = FAIR!

Open Science

Open Educational
Resources

Open Access

Open Peer Review

Open Methodology

Open Source



Open Data



TOUJOURS DE L'ESPOIR, IL Y A

Open data, motivations ou obligations ?



Visibility ?

ORD championship ?

Open Research Data national strategy and action plan ?

Institutional ORD policies ?

Publisher requirements ?

Funders requirements ?

FUNDERS EXAMPLES...

SNSF

- **Mandatory DMP** for grant applications (since Oct. 2017)
- Researchers must share the **data underlying their publications**

ERC

- **Mandatory DMP** that underline FAIR and ORD to obtain funding
- Research data are **open by default**

Quels sont les challenges RDM que vous rencontrez ???



Every day RDM: some challenges of EPFL researchers

I have written a first **draft of the DMP** [...] and I would like to kindly ask you if you can **review** it

[...] recommend a **code license** for a collaborator - they want open-source [...]

Can you help me with the **storage** options available? [...] I plan to generate about **10-50 TB** data

[...] help us draft of the **consent** documents for the interviewees to sign. Is there any kind of **template** [...]

[...] savoir si le stockage [...] est possible sur les plateformes **cloud** tels que Google Drive, Dropbox, Onedrive, etc.

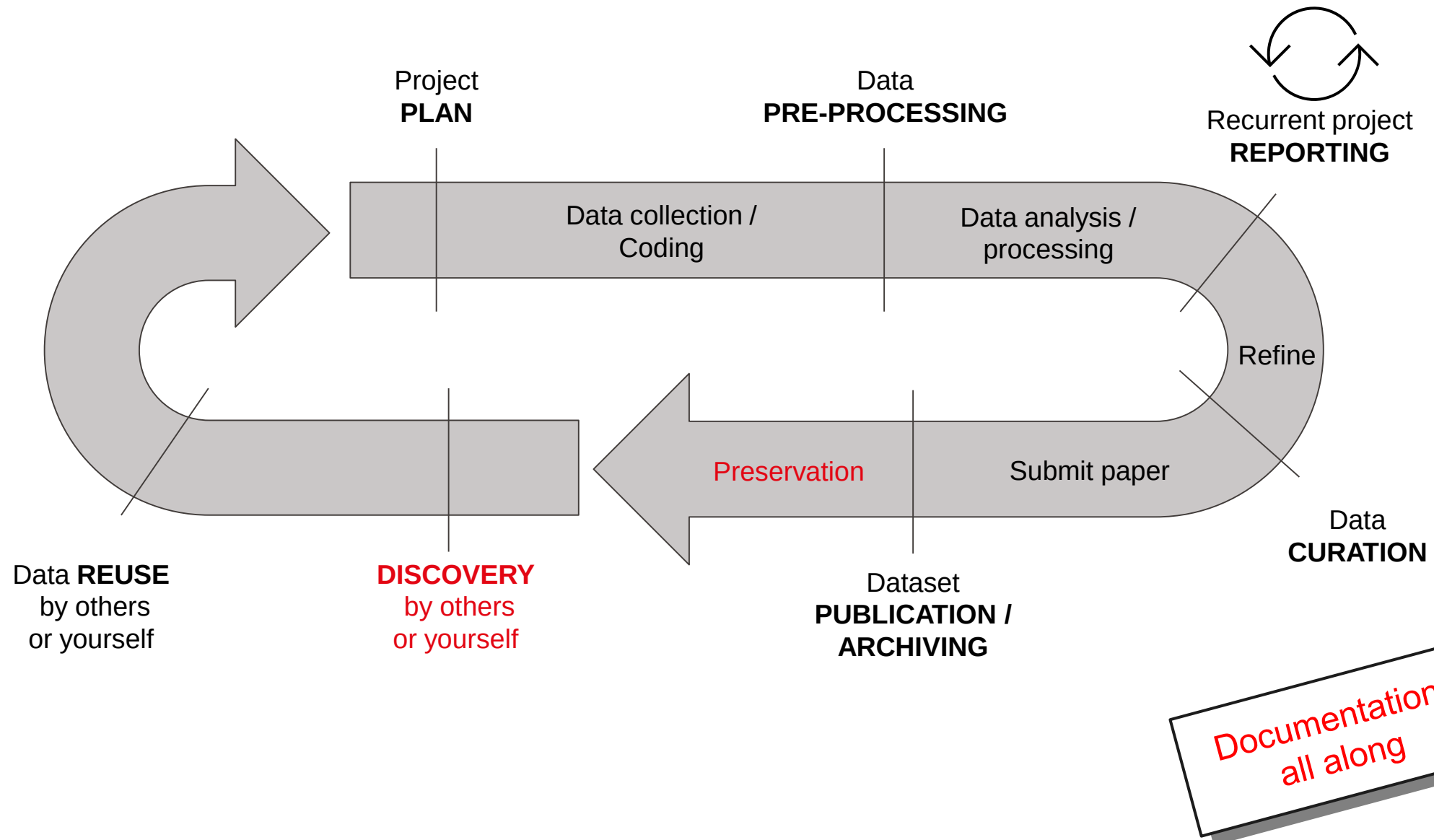
We have about **15 TB** of raw data [...] The **journal** is asking us to make this freely available [...]

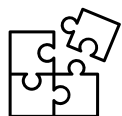
solution for managing the data [...] during the project, make it accessible to the **three labs** and make it eventually open source

I would like to use LINESURVEY for a data collection [...] would need to ensure the **privacy of the participants**

... etc. ...

Research Data Life Cycle (ideally)



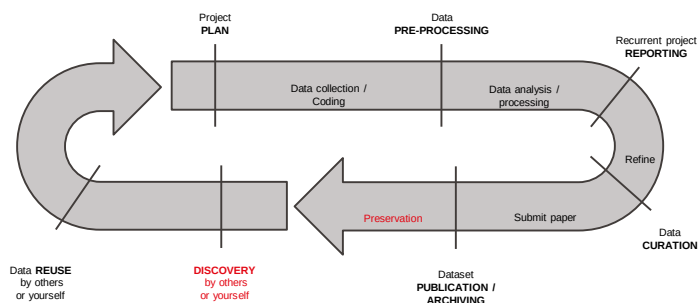


RDM Strategy (for > 1 projects, collaborators, groups)



DMP (for 1 project)

**Living document /
Research roadmap**



- **Anticipate:** future needs (material, software, HR ...)
- **Science:** impact, better reproducibility, posterity
- **Data reuse:** better use of public funds
- **Openness:** impact, transparency, accountability
- **Modernity:** world scale digital research, big data
- **Visibility:** citations, collaborations, career
- **Compliance:** law (ex. GDPR), funders (ex. SNSF)
- **Efficiency:** faster research for your lab and beyond
- **Risk reduction:** data loss, privacy, patents, ...

DMP Templates

Several templates of Data Management Plans (DMP) for different research funders, to ease your DMP submission and follow-up. The EPFL templates contain practical examples and recommendations.

[SNSF DMP \(with examples\)](#)

[SNSF DMP \(only guiding questions\)](#)

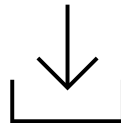
[ERC DMP](#)

[H2020/MSCA DMP](#)

[EPFL DMP](#)

[NCCR RDM Strategy \(series 5\)](#)

[EPFL RDM Strategy](#)



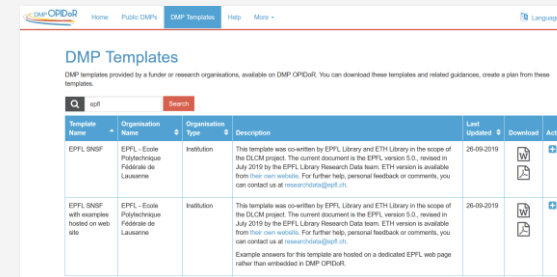
RDM walkthrough guide

From planning to preservation, use this [introductory guide](#) to Research Data Management, with links to online resources and practical suggestions.

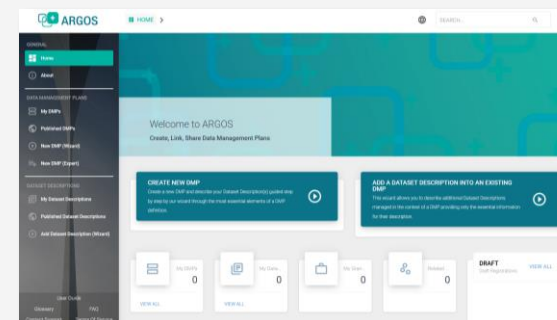


go.epfl.ch/rdm-guide

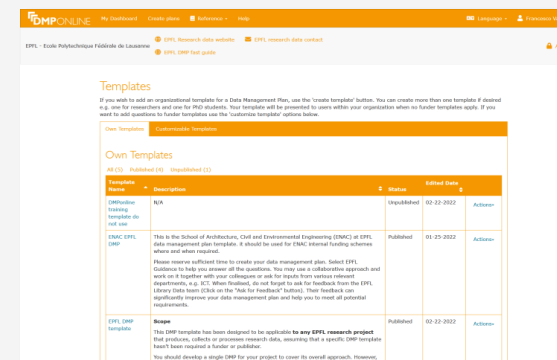
Other resources



dmp.opidor.fr/public_templates



argos.openaire.eu



dmponline.dcc.ac.uk

- **Data management planning** (ex. [DMPonline](#), [DMP OPIDoR](#), [Argos](#), [Storage Cost Calculator](#))
- **Code development / versioning** (ex. [GitLab](#), [GitHub](#))
- **Data processing / workflows** (ex. [Renku](#), [thot-data](#), [OpenRefine](#), [AiiDA](#))
- **ELNs / LIMS** (ex. [c6h6.org](#), [SLIMS](#), [ELN Comparison Matrix](#))
- **Legal issues** (ex. [DMLawTool](#), [Graasp Insights](#), [TLDRLegal](#)) ... plus tard... 😊
- **Specialized/Field-specific** (ex. dataviz, spettroscopy, simulation, surveying,...)
- **Data/Code dissemination** (ex. [OLOS](#), [Zenodo](#),... , [re3data](#), [comparative table](#), data journals)

Some good practices (I)

PLAN

DMP or RDM Strategy. Make it practical. Put links. Update it.

STORE

Use the 3-2-1 rule. Make sure coworkers have access.

ORGANIZE

Decide a file organization. Write it down. Share it with coworkers.

*ANONYMIZE

Secure raw data. Anonymize soon. Work on anonymized data.

CATEGORIZE

Separate raw/processed/final data and code. Excel is not a database

Some good practices (II)

CONVERT

Use open formats and open software from the beginning.

README

Make it “dumb”-proof. Target reproducibility.

PROTOCOL

Document methods and practical how-to.

CLEAN

Update README, protocols, plan,... Discard garbage.

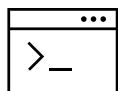
OUTPUT

Publish article + Code + Data + Protocols. Preserve/archive it ☺

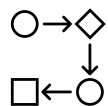
Towards publication: **FAIRify**



DOCUMENTATION Add context to make data easier to reuse, by yourself and others.
Ex.: **README** files, parameter files, data codebooks, etc.



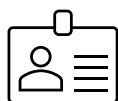
FILE FORMATS: Privilege (or convert to) open formats, make it future-proof.
Ex.: **CSV**, **PNG**, **ODT**, **FLAC**, **TIF**, ...



METADATA: Enrich your data to make them understandable by humans and machines.
Ex.: check **FAIRsharing** or use any published glossary.



ACCESS: Publish your research with a **link** to the openly access the underlying dataset.
Ex.: on publisher's website, institution's repository, other data repository.



PERSISTENT IDENTIFIERS: Attach a long-lasting reference to data and metadata.
Ex.: **DOI** (objects) and **ORCID** (authors).



DATA LICENCES: Specify in the publication what users can do with your data.
Ex.: **CC0** and **CC-BY-4.0**

SENSITIVE

- Tests on **animals / humans** (Swiss Animal Welfare Act, Human Research Act)
 - **Personal data** (Federal Act on Data Protection, GDPR)
HINT: Any privacy risk? Data protection measures to review? Are subjects well informed?
-

COMMERCIAL

- **3rd party data** (e.g. commercial datasets, research cooperations, etc.)
HINT: Check out the **contract** for data usage / sharing ... Or make one!
- **Potential commercial / industrial exploitation or patents**
HINT: Choose the **data license** + tell in the DMP!

Among other reasons... To avoid this!



Source: <https://twitter.com/robertnulrich/status/1490567709188182016>

- **Support around RDM**

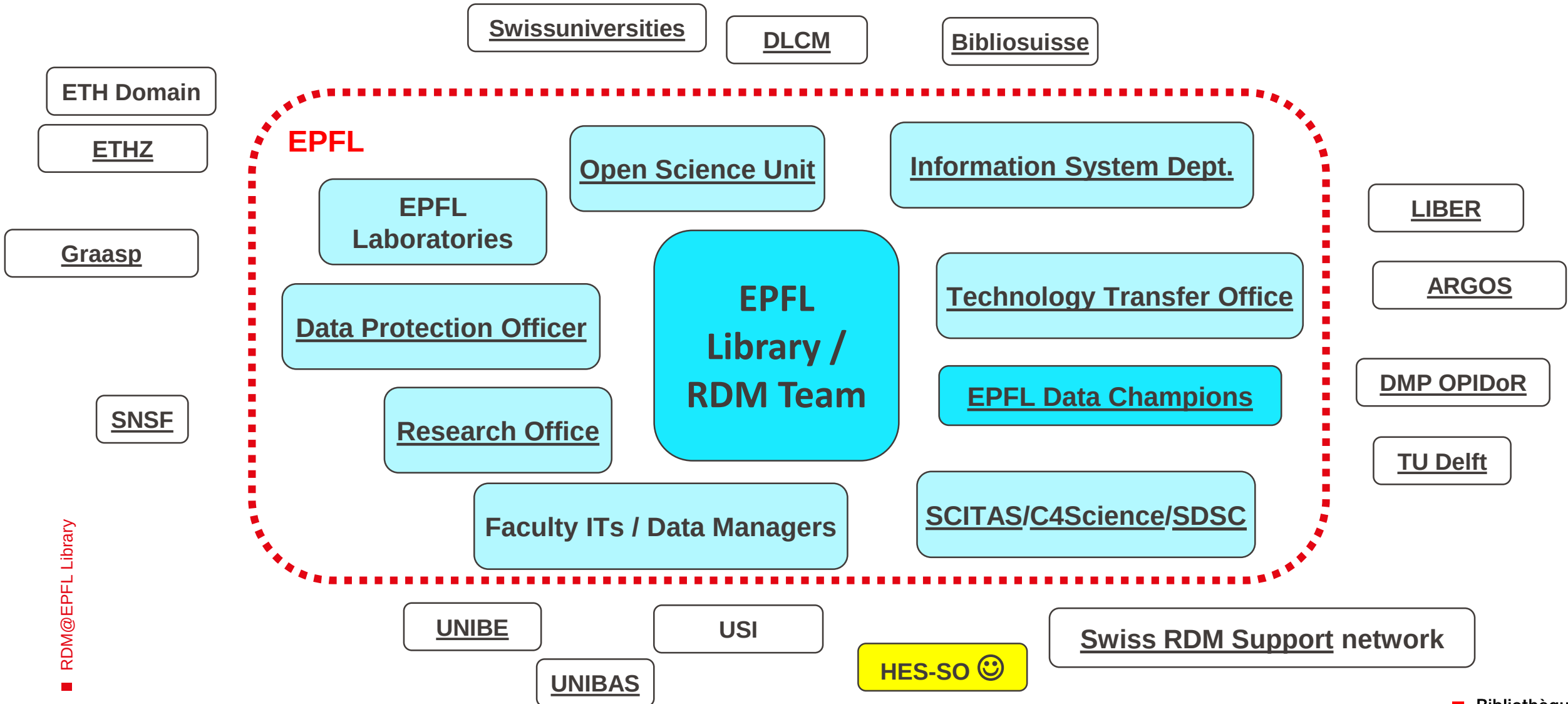
- Storage & Backup
- Data organization & Documentation
- Compliance with funders / laws
- RDM cost evaluation
- Publication platforms (i.e. repositories, data journals)
- Anonymization methods
- ELNs & other software
- Metadata and standards
- Licenses
- Data curation
- ...



- For researchers and students
- Aligned with policies (institutional, national, etc.)
- Partner of other univ. / HES services

L'importance des réseaux...

exemple de la Bibliothèque de l'EPFL



Chaque institution est différente



To find support, ask your local

- Library
- Data manager
- Open Science advisor
- Research Office
- ...

Questions





go.epfl.ch/rdm

go.epfl.ch/rdm-start

go.epfl.ch/rdm-guide

go.epfl.ch/rdm-software

go.epfl.ch/rdm-training

go.epfl.ch/rdm-support

go.epfl.ch/rdm-contact