

The background of the slide features a close-up photograph of a white ceramic coffee cup filled with dark coffee, sitting on a wooden desk. In the foreground, a spiral-bound notebook with lined pages is open, and a bright blue pen lies diagonally across it. A laptop is partially visible in the background, out of focus.

How to choose your data repository

In 10 minutes

**Francesco
Varrato**

EPFL Library

**Coffee Lecture
#07**

May 5, 2021

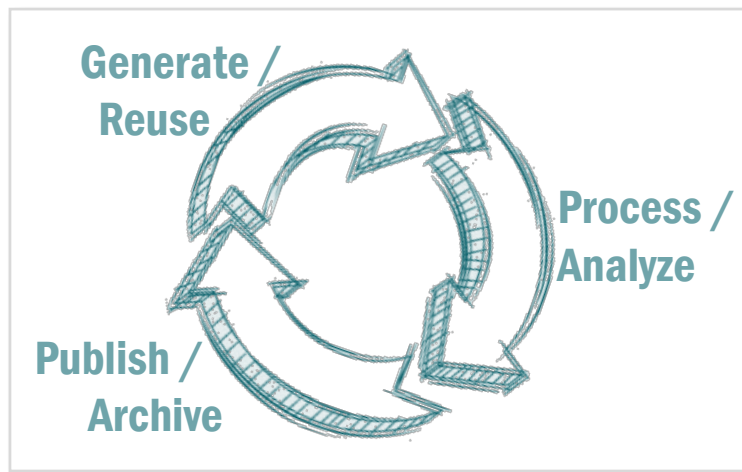
 = **where to start** looking for a data repository

 = **what to check** when selecting a data repository

📍 1/4 – Data Repository is ...

∞ Preserve^[1] and disseminate datasets

➡ Discover, retrieve and reuse datasets



... the best of **publishing, archiving & reusing**

 Most used^[2]: [Zenodo](#)

 Vertical search engine: [re3data](#)

GitHub or EPFL servers are **not data repositories**

📍 3/4 – Recommended



SNSF^[3]

- Dryad, EUDAT, OSF, Zenodo, ...



ERC^[4]

- Dataverse, Dryad, EUDAT, Figshare, OSF, Zenodo, ...
- + domain-specific ones

Depends on **who pays the bills**



Colleague(s)

- Group's data manager
- NCCR data official / Dept. scientific advisor / ...
- Local Data Champion^[5]



researchdata@epfl.ch



Eliane



Sitthida



Alain



Antoine



(myself)



Communities

Showing 0 to 10 out of 14 communities.

Sort by ▾

EPFL - École Polytechnique Fédérale de Lausanne

[View](#)

Curated by: rdmeplf

CHILI Lab, EPFL

[View](#)

Shared Resources for Computer Human Interaction for Learning and Instruction Lab at EPFL, Lausanne, Switzerland.

Curated by: StianHaklev

From Neuroscience to Machine Learning

[View](#)

Presentations of the Human Brain Project SP4 / CPD5 workshop "From Neuroscience to Machine Learning" held at the European Institute for Theoretical Neuroscience (EITN) at Paris. Workshop Abstract: The workshop aims to bring together researchers from...

Curated by:

Digital Object Identifier

e.g. 10.1234/foo.bar

[Reserve DOI](#)

Publication date *

Title *

Authors *

Family name, given names

Affiliation

[ORCID](#) (e.g.: 0000-0002-1825-0097)

[+ Add another author](#)

Description *

Rich text editor toolbar with icons for bold, italic, strikethrough, link, unlink, list, quote, code, undo, redo, italic, bold, link, unlink, source, and fullscreen.

Required.

Version

Keywords

Access right *

- ☒ [Open Access](#)
- ☐ [Embargoed Access](#)
- ☐ [Restricted Access](#)
- ☐ [Closed Access](#)

License *

Creative Commons Attribution 4.0 International

Grants

European Commission (EU)

[+ Add another grant](#)

Related identifiers

e.g. 10.1234/foobar.56789

N/A

Optional. Resource type of the related identifier.

Filter

Subjects 

Content Types 

Countries 

AID systems 

API 

Data access 

Data access restrictions 

Database access 

Database access restrictions 

Database licenses 

Data licenses 

Data upload 

Data upload restrictions 

Enhanced publication 

Institution responsibility type 

Institution type 

Keywords 

Metadata standards 

PID systems 

Provider types 

Quality management 

Repository languages 

Software 

Repository types 

Versioning 

Repository details

Materials Cloud Archive



General

Institutions

Terms

Standards

Name of repository

Materials Cloud Archive

Additional name(s)

Materials Cloud

Repository URL

<https://archive.materialscloud.org/>

Subject(s)

Engineering Sciences Materials Science and Engineering Condensed Matter Physics Natural Sciences

Chemical Solid State and Surface Research Physics Chemistry Theoretical Condensed Matter Physics

Theory and Modelling Materials Science

Description

FAIR & long-term storage of research data from computational materials science, with particular focus on sharing the full provenance of calculations.

Contact

giovanni.pizzi@epfl.ch

nicola.marzari@epfl.ch

Content type(s)

Raw data Scientific and statistical data formats Databases

Keyword(s)

physics atomistic simulations material science materials informatics computational materials science FAIR

Persistent identifier(s) of the repository

FAIRsharing_doi:10.25504/FAIRsharing.tlbUpj

EPFL 3/4 – Comparison

9/10

Francesco Varrato

NAME	DISCIPLINE	NOPROF / INST.	COUNTRY	FREE	MAX VOLUME	LICENSING
	Generic	✓ (CERN)		✓	50GB/dataset, ∞ datasets	CC, GNU, BSD
 MATERIALSCLOUD	STI / Materials	✓ (EPFL)		✓	5GB General, 50GB AiiDa DB	CC-BY (MIT for AiiDa)
	Generic	✗ (Holtzbrinck Group)		Freemium	1 TB per dataset	CC0, CC-BY
	Bio / Medical	✓ (?)		✗	?	CC0
	Generic	✓ (Harvard University)		✓	?	?
	Generic	✓ (HORIZON 2020)		✓	?	CC (DARUP)
	Aquatic	✓ (Eawag)		✓	Unlimited upload size	CC0 default / Changeable
	Humanitarian Data	✓ (United Nations)	International	Only approved organisations	?	CC BY-IGO, CC BY, CC BY-SA, ODC-ODbL, ODC-BY, PDDL, ...

Ongoing: comparative table w/ ENAC [Data Manager](#) & STI [Scientific Advisor](#)



0. Preserve your datasets! ... but not all
1. Target public: field-specific vs. generic (both?)
2. Listed on re3data: peace of mind
3. DOI or other Persistent Identifier
4. Cross-linking: dataset / code ↔ article
5. Non-profit: ex. SNSF doesn't reimburse
6. Good licences choice: reuse & compliancy
7. Max upload: size matters

Depends on **needs, strategy, time**



go.epfl.ch/coffee-lectures



go.epfl.ch/rdm

- [1] CASRAI, *RDM Glossary* - Preservation
- [2] SNSF (2020), *Open Research Data: SNSF monitoring report 2017-2018*
- [3] SNSF (2020), *Which data repositories can be used?*
- [4] ERC Scientific Council (2019), *Open Research Data and Data Management Plans*
- [5] EPFL Data Champions community webpage
- [6] Husen & al. (2017), *Recommended versus Certified Repositories: Mind the Gap*
- [7] Science Europe (2021), *Practical Guide to the International Alignment of RDM*, p.26

Contacts & credits



francesco.varrato@epfl.ch



researchdata@epfl.ch



2021 Francesco Varrato.

This is an open-access document distributed under the terms of the Creative Commons Attribution License