
150 ans
Société Française
de Physique
150 ANS D'ENGAGEMENT
POUR LA PHYSIQUE

How to share your research outputs?

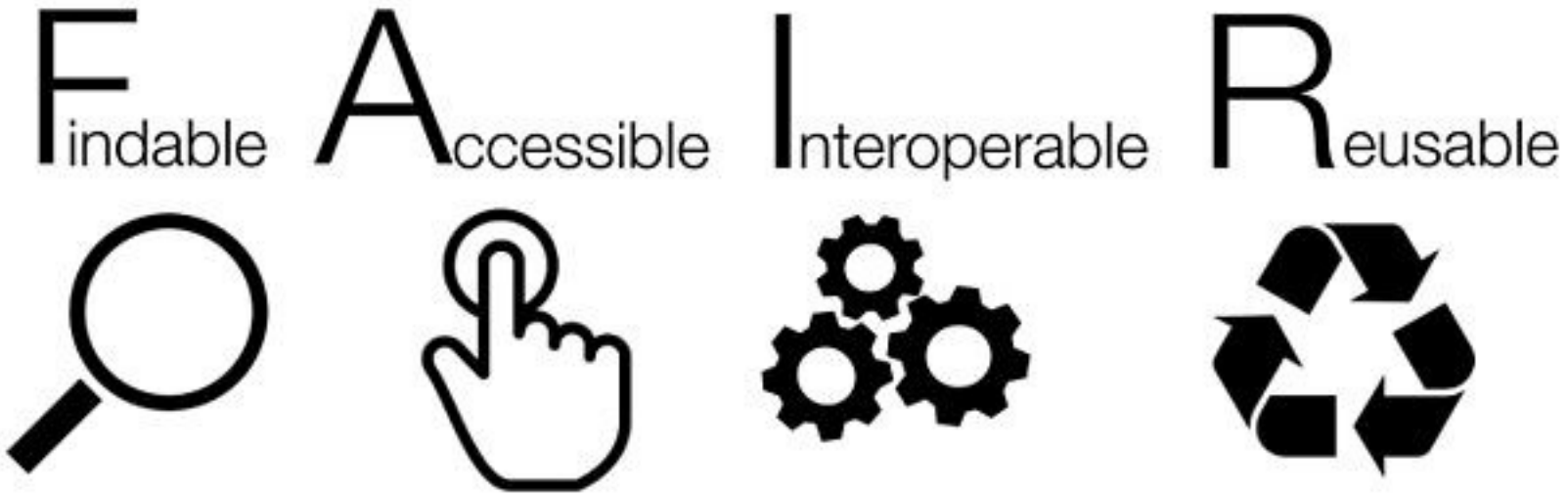
Service des **bibliothèques**
Université de Strasbourg

Stéphanie Cheviron

20 june 2023

Sharing research outputs needs to be **anticipated in order to promote its reuse**. This means paying particular attention to outputs management throughout its life cycle, and applying the FAIR principles (findable, accessible, interoperable, reusable).

Research outputs: data, code, algorithms, protocols, ELN, etc.



https://www.go-fair.org/fair-principles/

The FAIR principles

There is no **single, one-size-fits-all way to manage research data and make them FAIR**. What is appropriate and feasible largely depends on the research domain and data type(s) involved, as well as on the specificities of the project.

FAIR does not necessarily imply Open; data can be FAIR and shared under restrictions.

Source: https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm

Code citation

```yaml
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/1234-5678-9101-1121
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2021-08-11
```

Citation CFF file

https://citation-file-format.github.io/

Plain text files with **human- and machine-readable citation information** for software (and datasets). Code developers can include them in their repositories to let others know how **to correctly cite their software**.

Generate your own CFF file here:
https://citation-file-format.github.io/cff-initializer-javascript

Code preservation

Software Heritage

The universal software archive harvests GitHub, GitLab,… You can also deposit your code

https://www.softwareheritage.org/

HAL open science

You can deposit your code on HAL.
→upload a single .zip or .tar.gz archive.

How to deposit software source code on HAL?

Your code can be deposited on SH by HAL

Electronic lab notebooks

ELN improve data sharing and preservation. Researchers can store and share their (successful or unsuccessful) experiments and make them available more easily and rapidly.

- eLabFTW (chosen by CNRS)
  - Best practices (in French)

- Chemotion (KIT)
- Jupyter Notebooks and JupyterLab

More about ELN: https://www.datacc.org/en/best-practices/how-to-use-an-electronic-laboratory-notebook/

Protocols

Bring structure to your research

A secure platform for developing and sharing reproducible methods.

https://www.protocols.io/

Trusted Data repositories

- Deposit your data on a certified data repository
- Or a trusted repository:
  - is in open access
  - Delivers PID, usually DOI
  - Provides rich metadata
  - is moderated or is recognised by a scientific community
  - Has a management system for licenses and embargos

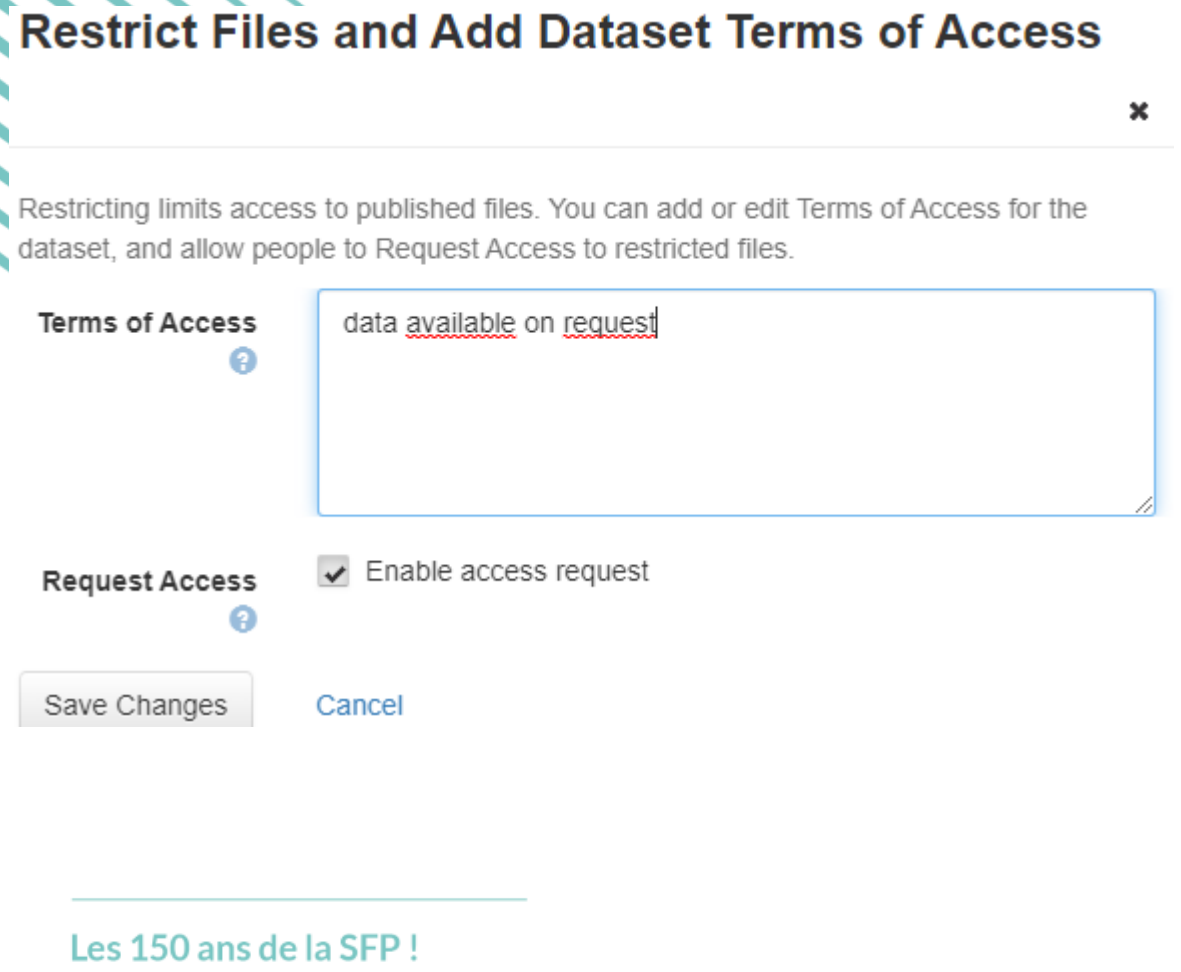
→ Check the data policy of the publisher
→ Deposit your data in a open access repository instead of supplementary materials on the publisher's platform

To find a repository (and data):

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES


AS OPEN AS POSSIBLE, AS CLOSE AS NECESSARY
THE RESEARCHER
imgflip.com

What if I can’t share data?



“*As open as possible, as closed as necessary*”

The dissemination of research data **does not mean opening it up to all publics**.

While unrestricted openness may be the chosen solution, it is only one of many. In addition to the legal impossibility of sharing of certain data, there are many reasons why justify the decision to restrict access: confidentiality issues for respondents, the need to produce primary analysis results first, etc.

You can use **embargo**, for example in *Recherche Data Gouv*:

https://recherche.data.gouv.fr/en/category/9/guide/depositing-a-dataset#6.3

File organisation

- File organization: by project, by analysis, by date, etc
- File naming convention
- Use **open file format** for better readability
- Use a standard for dates: ie. YYYY-MM-DD

More about naming conventions: https://www.datacc.org/en/best-practices/establishing-data-management-plan/naming-files-managing-versions-good-habits/

Quality and documentation

DOCUMENT YOUR DATA: without context, they can't be reuse

- Add a Readme.txt →https://www.makeareadme.com/
- Use **metadata standards** (Dublin Core, Datacite → generic or specific to your field): on a repository, in the dataset
- Write (and share) the **codebook**: explains all the concepts and/or variables that are present in the datafiles. A good codebook needs to be clear enough for a reader to be able to interpret and (re)use the data.

More information: https://www.ru.nl/rdm/vm/example-codebook/

How to write a good codebook: https://www.medicine.mcgill.ca/epidemiology/joseph/pbelisle/CodebookCookbook/CodebookCookbook.pdf

Data storage and safety

Storage must :

- securely store data on a medium
- allow collaborative access to this resource in the short or medium term
- Guarantee data integrity by protecting it from the risk of loss

Examples:

- Use the 3-2-1 Rule: “*don’t put all of your eggs in one basket*”
- Use an institutional cloud, not a commercial one
- Prepare a strategy for the long-term preservation of your data: what data will be preserved? Where? How long? → cold storage is costly

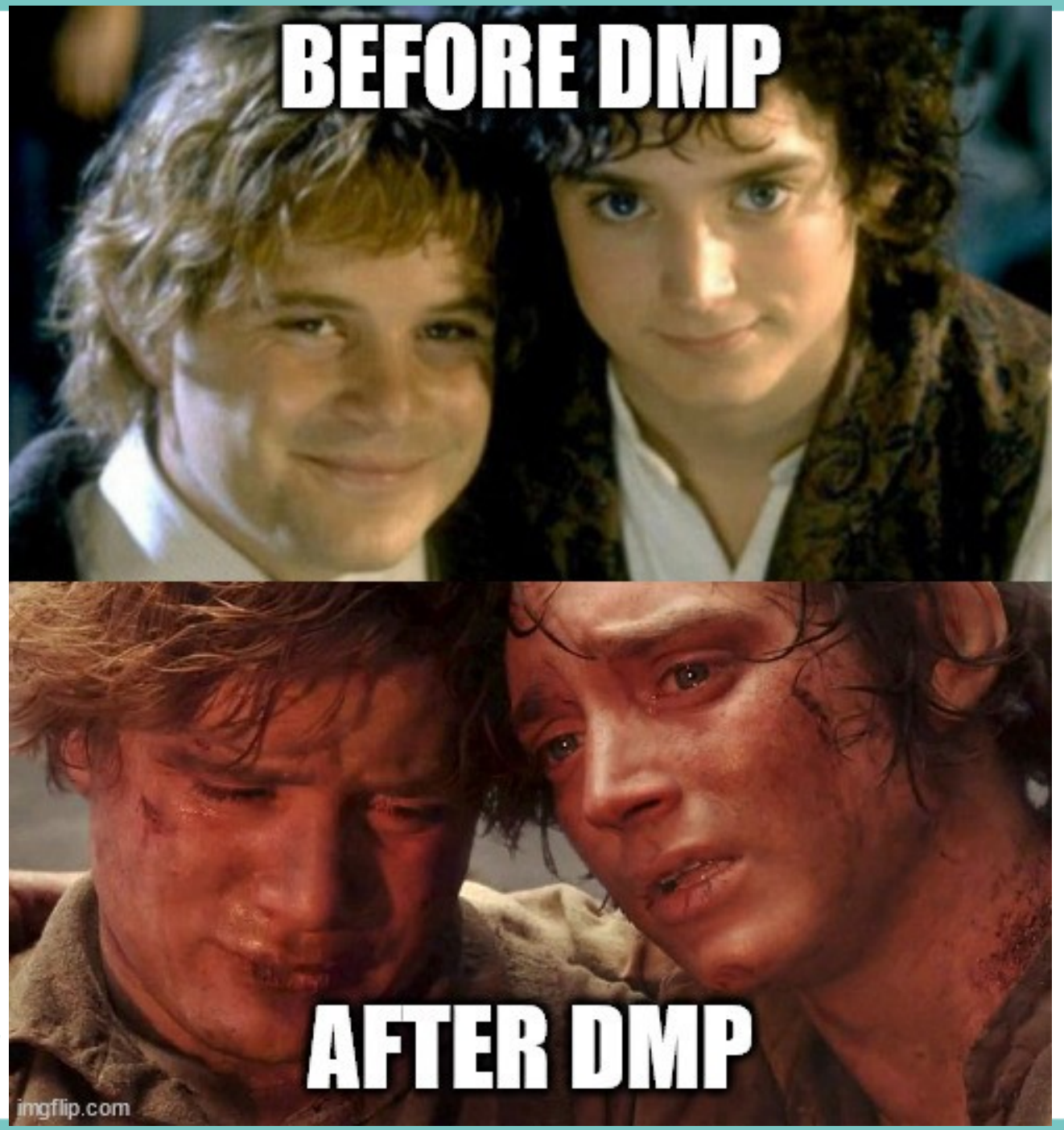
The Data Management Plan

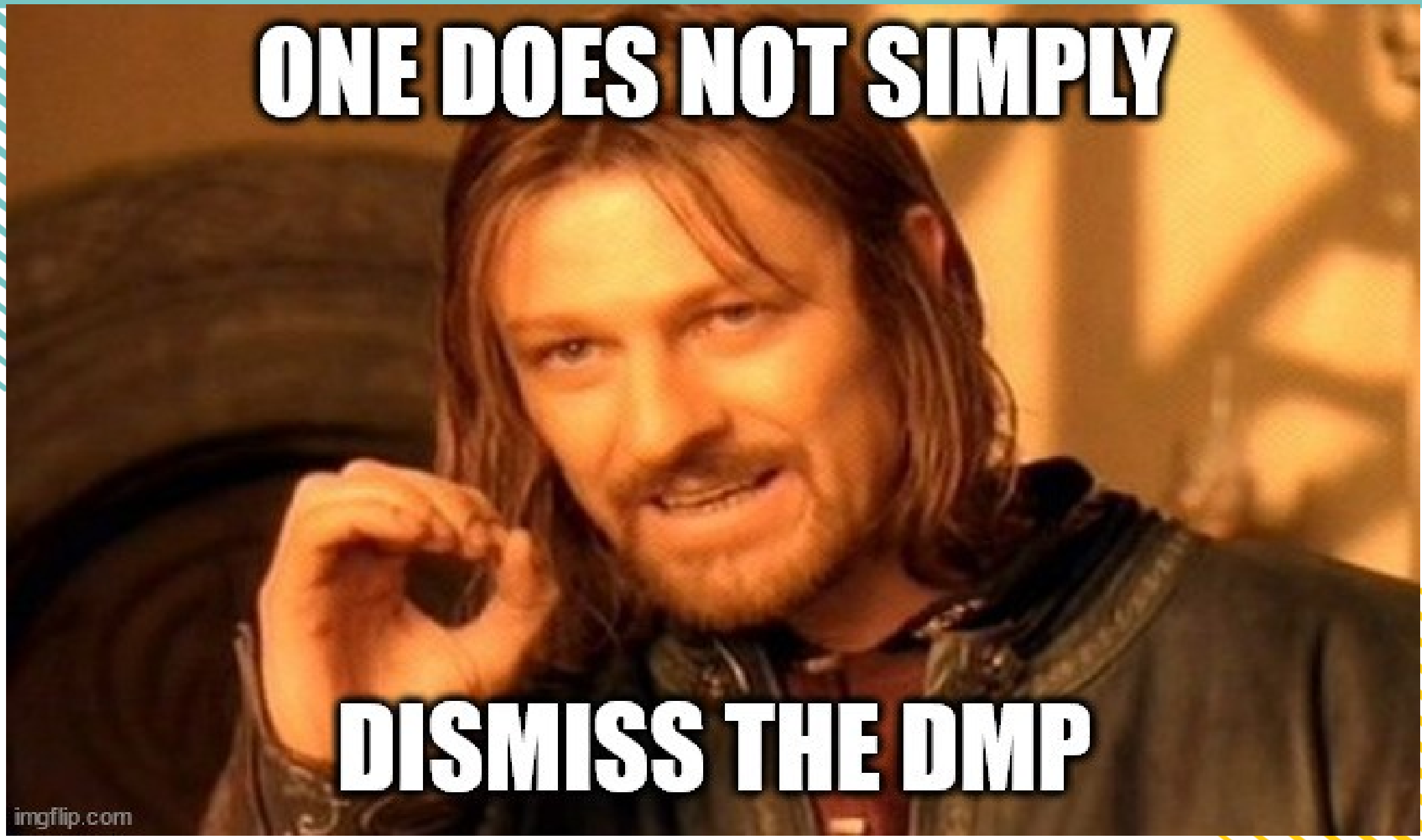
A DMP is a document that outlines, from the start of the project, how research data will be handled both during and after a research project.

It identifies **key actions and strategies** to ensure that research data are of a **high-quality, secure, sustainable**, and – to the extent possible – **accessible and reusable**.

Source: https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm


FUNDERS
MOUNTAIN
OF RED TAPE


BEFORE DMP
AFTER DMP
imgflip.com


ONE DOES NOT SIMPLY
DISMISS THE DMP
imgflip.com

- It’s a **planification tool for project management** used to estimate RDM, storage and preservation costs and write a better grant application (useful too for consortium agreement)
- During the research project, it’s a **living document** you update regularly, it is a reference.
- It's the combination of all the knowledge about your project data: it gives **context** to your data
- Important for:

→ reuse of data by other researchers

→ legal team, archivists, IT, library and every staff who will process the data

→A DMP =\= open science: even if the data can’t be shared, you must do a DMP: it’s good practice

Costs of RDM

Data management and sharing activities need to be costed into research, in terms of the time and resources needed. By planning early, costs can be significantly reduced. Costs associated with open access to research data, can be claimed as **eligible costs** of any Horizon Europe grant **during the duration of the project** under the conditions defined in the Grant Agreement: they must **already be budgeted and accepted in the grant proposal**, and note the "during the duration of the project".

To estimate costs for RDM, you can check the online RDM-costing tool and the infographic on 'What will it cost t o manage and share my data?'.

Source: https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm

Examples of RDM activities

- Curating data
- Developing supporting documentation
- Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
- De-identifying data
- Preparing metadata to foster discoverability, interpretation, and reuse
- Local data management considerations, such as unique and specialized information infrastructure necessary to provide local management and preservation (for example, before deposit into an established repository).
- Preserving and sharing data through established repositories, such as data deposit fees.

Source: https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/budgeting-for-data-management-sharing#after

Tools for writing a DMP

https://dmp.opidor.fr/

Templates of funders, universities, research institutes + **software management plans** (SMP)

https://argos.openaire.eu/

Guidances about DMPs

- Guiding you in Open Science, OpenAIRE
- *General Model Grant Agreement* , annexe 5 (article 17), v.1.1, 15/04/2022 [PDF]
- Annotated model grant agreement, annexe 5 (article 17), v.0.2, 30/11/2021 [PDF]
- *Programme Guide Horizon Europe*, v.2, 11/04/2022 [PDF]
- *Modèle de DMP Horizon Europe*, v.1, 05/05/2021 [DOCX]
- Open Open reasearch data and data management plans information for ERC grantees , v.4.1, 20/04/2022 [PDF]
- webinaire OpenAIRE Horizon Europe Open Science requirements in practice, SLIDES [PDF]

Guidances about DMPs

- « Grille de relecture » for the ANR DMP template ← use it as a checklist
- Practical Guide to the International Alignment of Research Data Management – Extended Edition, Science Europe, janvier 2021

Guidance about RDM

- The Open Science Training Handbook
- DATACC for physics and chemistry
- RDMkit for life sciences (Elixir)

NIH
ANR (in french)
Wellcome Trust (Outputs management plan)

CNRS Data policy

- Plan données de la recherche du CNRS
- CNRS : un plan ambitieux pour des données accessibles et réutilisables
- ESCAPE : les physiciens préparent leur science ouverte
- Fiche Repère CNRS « la diffusion des données » [PDF, in french]


THE
FELLOWSHIP
OF THE
RING
DMP


YOU'RE NOT ALONE
SECURITY OFFICER
ARCHIVIST
LEGAL EXPERT
IT EXPERT
RESEARCH OFFICER
RESEARCHER
DATA LIBRARIAN
DATA OFFICER
PLATFORM MANAGER

Merci

150 ans

150 ANS D'ENGAGEMENT
POUR LA PHYSIQUE

Stéphanie Cheviron

scheviron@unistra.fr
bu-data@unistra.fr
contact@adele.help