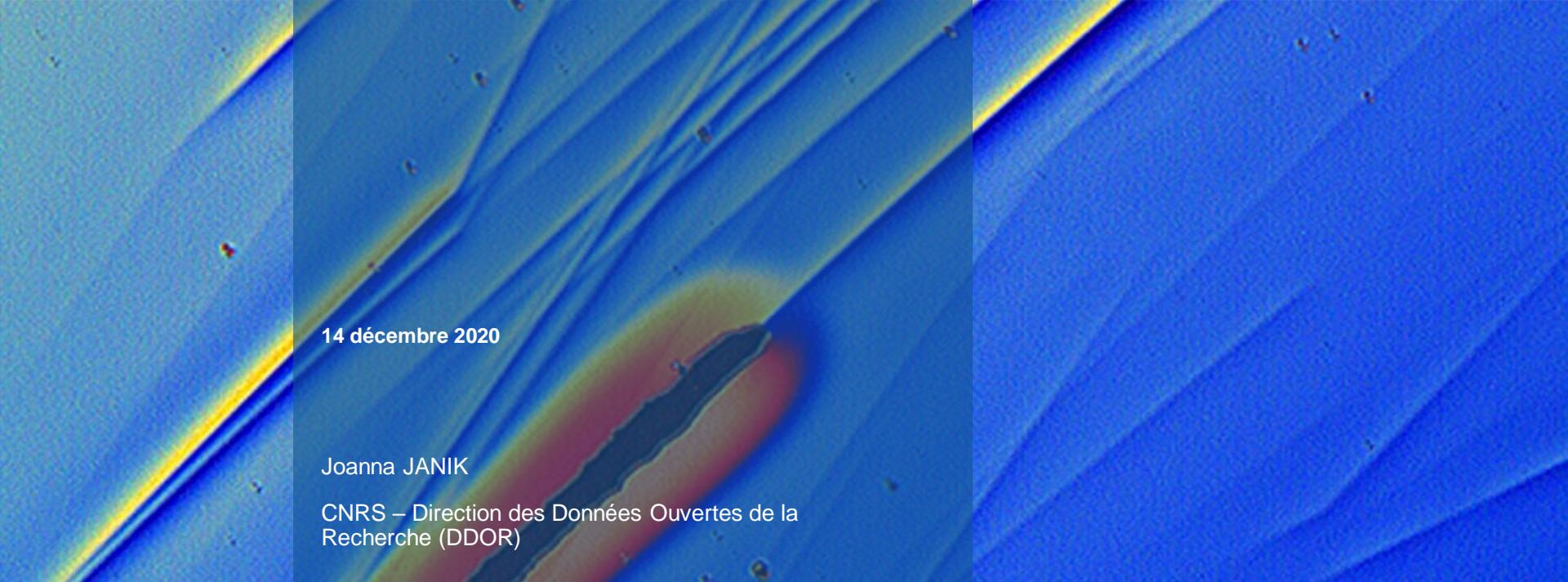


« Opening Science » - LabeX UnivEarthS Thematic School 2020

Open science: how to begin ?

A wide-angle aerial photograph of a coastal landscape. A prominent yellow line runs diagonally across the frame, starting from the top left and ending at the bottom right. The terrain is a mix of green vegetation and blue water. In the bottom left corner, there is a small white text overlay.

14 décembre 2020

Joanna JANIK

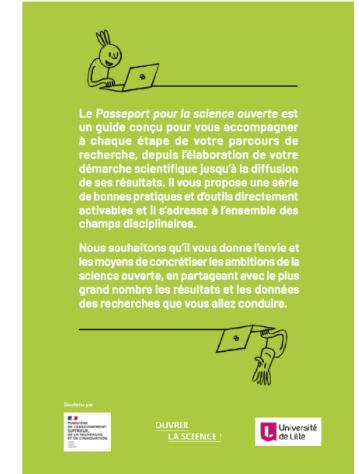
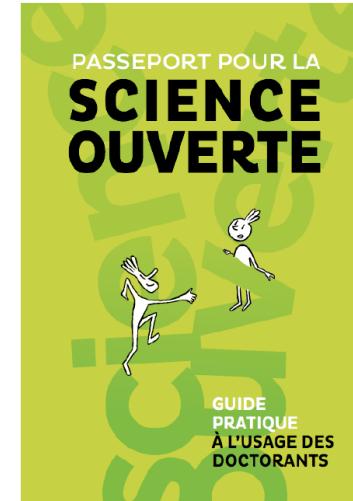
CNRS – Direction des Données Ouvertes de la
Recherche (DDOR)

Agenda

- Why Open Science?
- Disseminate your research
 - Writing and disseminating my thesis
 - Disseminating your publications in open access
 - Producing and managing research data
 - (some) Legal issues
- Focus
 - Reproducibility
 - Codes and softwares open source
 - Research evaluation
 - Text and data mining
- Going further

Le Passeport : livret illustré de 40 pages

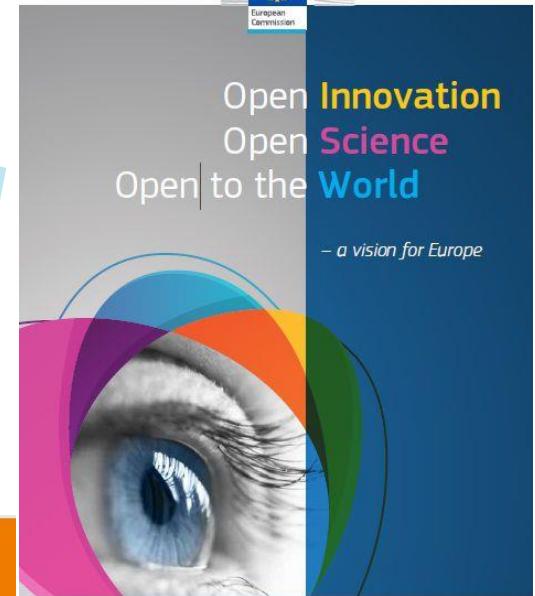
Couverture et 4ème de couverture



Why Open Science?

- Principle : Public money ?
Public data !
- Scientific : better quality
research
- Society : a more open science,
more accessible to society

*« The goal of OS is to use
and reuse your results »*
SHB online



What?

FOSTER defines Open Science (OS) as 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.

Why?

Opening the research process supports validation, reproducibility and reduces cases of academic misconduct. It helps to maximise the impact of your research and provides the foundations for others to build upon. In short, applying open science in your daily workflows is just part of good research practice!

<https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>

Diffuser sa recherche

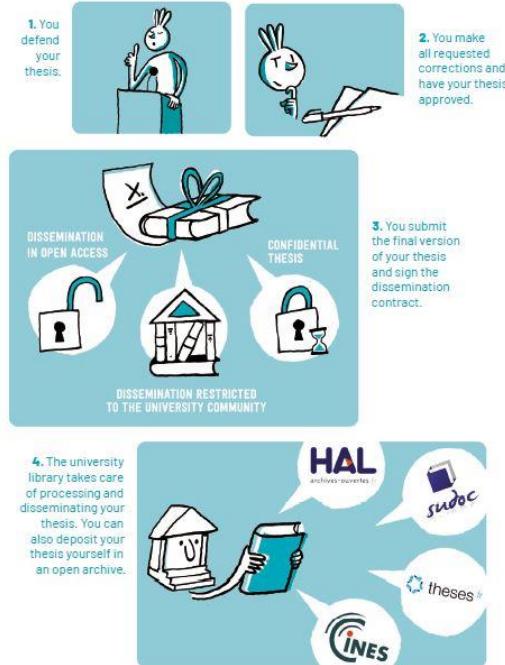


Je rédige ma thèse et je la diffuse

JE SUIS DOCTORANT EN CHIMIE : POUR MOI, CE QUI COMpte, CE SONT PLUS LES RÉSULTATS D'EXPÉRIENCE QUE LA RÉDACTION

- 👉 Quels outils ouverts me permettraient facilement de retrouver des références que je mentionne et de respecter les normes bibliographiques ?
- 👉 Comment rédiger un document qui garderait sa structuration/sa mise en forme pour qu'il soit lu sur différents supports numériques ?
- 👉 Comment permettre l'accès aux résultats de mes recherches au plus grand nombre ?
- 👉 Comment archiver de manière sûre les résultats de mes recherches mais aussi les données liées ?

The stages of depositing and disseminating



Disseminate your research

1 – I write my thesis and disseminate it

- Write in accordance with the principles of OS

- open collaborative tools
 - bibliographic referencing tools
 - editorial tools
 - open source softwares

- Disseminate your thesis in open access
- Create a researcher ID – ORCID



<https://orcid.org/>



II. Disseminate your research

2- I disseminate my publications in open acces

- Open archives
(exemples: [HAL](#), [ArXiv](#), [BiorXiv](#), [RePEc..](#))
 - Deposit on preprints servers
- Open access journals (directories [DOAJ](#), [Bioline](#), [Electronic Journals Library...](#))
 - financial conditions (hybrid reviews, with APCs)
 - « predatory » journals - cf [Stop predatory journals](#)



Disseminate my research



Je produis des données

JE SUIS DOCTORANT EN HISTOIRE : JE NE PRODUIS PAS DE DONNÉES, MAIS J'EN RÉCOLTE BEAUCOUP

- 👉 Quels outils utiliser pour bien gérer mes données ?
- 👉 Quelles sont les bonnes habitudes à prendre pour produire des données de meilleure qualité ?
- 👉 Comment faire un plan de gestion de données ?
- 👉 Comment puis-je procéder pour rendre mes résultats reproductibles ?
- 👉 Comment partager mes données tout en montrant que c'est moi qui les ai produites en premier ?

How to disseminate your data well

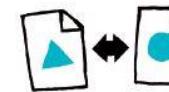
RESPECT THE FAIR PRINCIPLES



The aim of the **Findable** principle's is to facilitate the discovery of data by humans and computer systems and requires the description and indexing of data and metadata.



The **Accessible** principle encourages the long-term storage of data and metadata and facilitating their access and/or downloading by specifying the conditions of access (open or restricted) and use (license).



The **Interoperable** principle can be broken down as follows – data should be downloadable, usable, intelligible and combinable with other data by humans and machines.



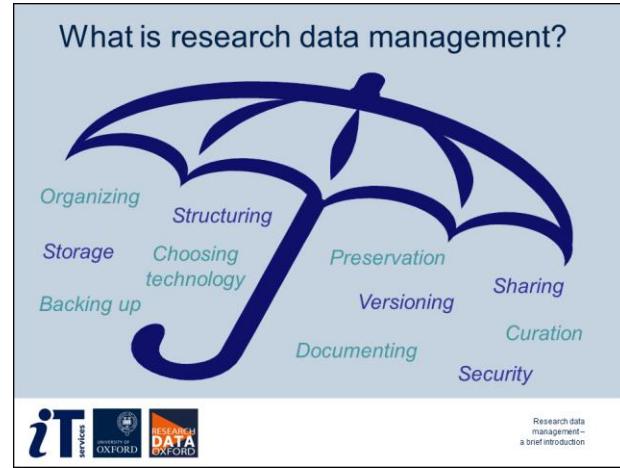
The **Reusable** principle highlights the characteristics that make data reusable for future research or other purposes (education, innovation, reproduction/transparency of science).

Source: *Produire des données FAIR*. In Ozalé Yeumo, E., L'Hostis, D., Coaud, S. Gestion et partage des données scientifiques. INRA [Internet]. 9/08/2018 [consulted on May 15th 2020]. Available at <https://www6.inrae.fr/datapartage/Produire-des-donnees-FAIR>

Disseminate my research

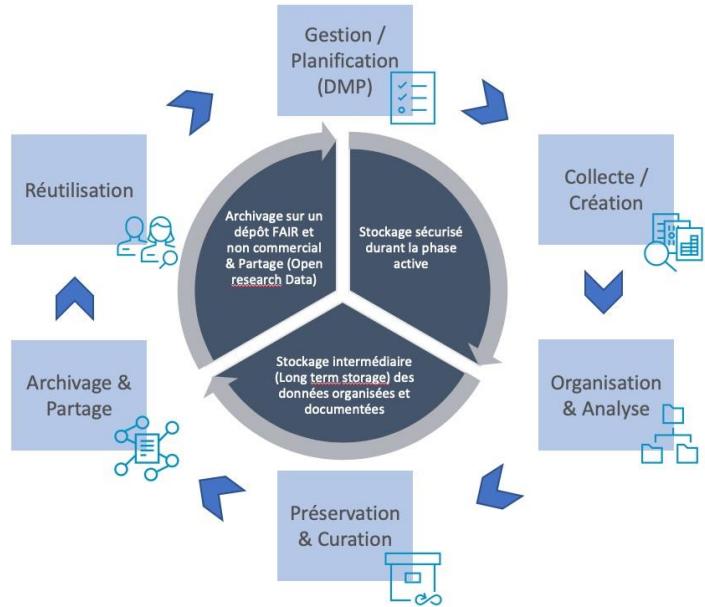
3 - I produce and manage research data

- Why manage research data (RDM) ?
 - Visibility, accessibility, traceability
- Who is concerned?
 - Researcher/Phd – (IT service) network administrator, developer – librarian (archivist) – lawyer..
- Data lifecycle -> research project



Acting throughout the life cycle of research projects

Cycle de vie des données de recherche



LE CYCLE DE VIE DES DONNÉES

Intégrer la gestion de données scientifiques aux activités de recherche

Le cycle de vie des données scientifiques est l'ensemble des étapes de gestion, de conservation, de diffusion et de réutilisation des données scientifiques liées aux activités de recherche.



Disseminate my research

4 - I produce and manage research data



- Integrate the Data Management Plan (DMP) into your research project
- All H2020 and ANR projects must provide the DMP

- How to be FAIR ?

Findable – Accessible – Interoperable – Reusable

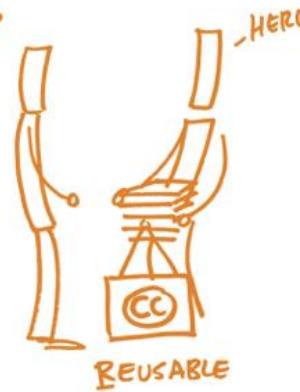
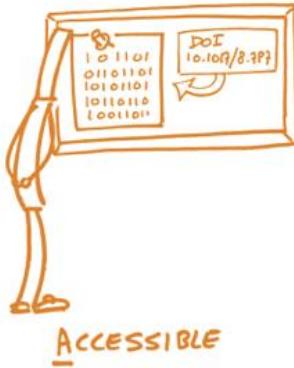
Guidelines on FAIR Data Management in Horizon 2020: Version 3.0.

European Commission, 26 July 2016



FAIR DATA PRINCIPLES

AH!



How-to-make-your-data-fair
Openaire.eu

Wilkinson, M. D., Dumontier, M. et al. (2016)

The FAIR Guiding Principles for scientific data management and stewardship
in *Scientific Data*

Disseminate my research

5 – Legal issues



■ Why open the data?

Loi pour une République numérique
octobre 2016

■ Promote the open access licenses

Creative Commons

■ « Je publie quels sont mes droits ? »

Comité pour la Science ouverte, 2^e éd., octobre 2020

Je publie,
quels sont mes droits ?

en 35 conseils

SOMMAIRE

COMPRENDRE :

Quel est l'impact du droit d'auteur sur mon travail de chercheur ? p.3
Conseils 1 à 7

FOCUS :

La science ouverte, c'est pour moi ! p.7
Conseils 8 à 12

PENDANT LA RÉDACTION :

J'identifie sur quoi porte mon droit d'auteur p.12
Conseils 13 à 16

PENDANT LA RÉDACTION :

Je suis prudent à l'égard des contenus extérieurs que je réutilise dans mon travail..... p.15
Conseils 17 à 24

FOCUS :

L'utilisation des images p.19
Conseils 25 à 28

LE MOMENT DE LA PUBLICATION :

Je vérifie le contrat d'édition p.22
Conseils 29 à 32

APRÈS LA PUBLICATION

Je reste vigilant p.26
Conseils 33 à 35

III FOCUS – Reproducible research

- Reproducible research – a prerequisite
- Reproducible research is not an end in itself
- Improving practices
- The relays
 - data processing and analysis engineers
 - statisticians



« Vers une recherche reproductible » HAL Id : hal-02144142, version 3
Desquillet, L., Granger, S. et al.

Software Heritage → collect – preserve – share

- **Archive and reference source code with save code now**
Archiving and Referencing Source Code with Software Heritage
- **Cite software in articles with pointers to SWH** (LaTeX users)
<https://www.softwareheritage.org/save-and-reference-research-software/?lang=fr>
- **Deposit in HAL and SWH**
Create software deposit: User guide and best practices

« Construire le pilier logiciel de la Science ouverte »

Software Heritage : « une infrastructure partagée, dédiée à la recherche, à l'industrie et au patrimoine culturel »

Le CNRS rejoint Software Heritage et apporte un soutien de 100 000 euros par an à cette bibliothèque universelle de codes sources de logiciels, lancée par Inria et soutenue par l'UNESCO. Son directeur, Roberto Di Cosmo en détaille les ambitions.

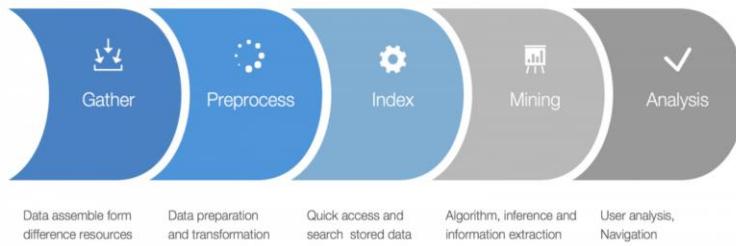
[Lire l'article](#)

III FOCUS – Text and data mining (TDM)

openMIN7ED Text Mining

Open Mining Infrastructure for Text & Data

Text mining involves a series of activities to be performed in order to efficiently mine the information. These activities are:



Source : free online course on text and data mining for 'non-tech people'.

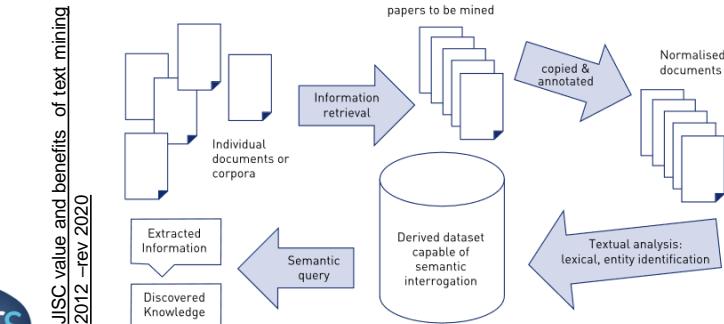


Figure 2 Schematic overview of the processes involved in text mining of scholarly content

TM - the set of methods and computer processing that consist in analyzing the meaning of texts in natural language in order to give a representation that can be used by both humans and computers

- ▼ Step 1: Is it protected?
- ▼ Step 2: What am I going to do with it?
- ▼ Step 3: For what purpose?
- ▼ Step 4: Do I have or need a licence?
- ▼ Step 5: Do I need further legal assistance?

Legal Guidelines for TDM practitioners



Transposing the Directive (03-2018) on Copyright in the Digital Single Market

Article 3. The new rules offer clear permission for libraries and their users to carry out text and data mining on all legally accessed materials, either on-site or remotely, including through put private partnerships. There are no further restrictions intended to ensure security of database or particular rules around the safe storage of datasets.

Article 4. All other users also benefit from a broad exception, with the possibility for rightholders to opt out, strictly defined and easy for users and machines to understand and respect.



What Open Science can bring:

- **It is the results themselves that must be evaluated**, not the fact that they may have been published in a prestigious journal or other reputable media.
- **For each of the productions cited** in the evaluation files, **researchers must explain the scope, impact, and personal contribution** they have made to the project.
- **All types of production** must be able to be part of the evaluation.
- All the productions quoted in the files of evaluation must be accessible in HAL or possibly in another open archive*.

* Too recent results may be embargoed. In which case they must still have been filed in HAL, with a duration of embargo not exceeding those provided by law (6 months in STM, and 12 months for SHS). They are then provided by a private link in HAL (or in the file).

IV Going further

- **Research Data Alliance (French national node)**
 - **The Foster portal** Open Science Training Courses
 - **CoopIST** (Cirad web site)
 - **Cat OPIDoR** (catalogue des services français dédiés aux données scientifiques)
 - **DoRANum** (dispositif des formations à distance)
-
- **GT « Atelier Données » du CNRS**
 - **Réseau des Urfist** (recherche et formation)
 - **SciencesPo Guide sur la SO**

Online soon : French version
of Foster Open Science
Training Handbook



« Opening Science » – LabeX UnivEarthS Thematic School 2020

and as always:

Site du CoSO

Site de Couperin

OUVRIR
LA SCIENCE

**Thank you for your attention
Questions ?**



Joanna.Janik@cnrs-dir.fr

