UNIVERSITÉ DE MONTPELLIER

Open Science

Open Science Seminar

February 7th, 2022

Lab1 - Master 1 Health Biology

Laure Lefrançois and Odile Hennaut
Research Support Service of the Office of Library Services
of the University of Montpellier

PLAN

1. What is Open Science ?
2. Open Access
3. Research Data
4. UM and Open Science

OPEN SCIENCE : WHATs & WHYs

This introductory course will help you to understand what Open Science is and why it is something you should care about. At the end of this session, you will

- understand what Open Science means and why it is important for you
- be aware of some of the different ways to make your own research more open over the research lifecycle
- understand why funding bodies are in support of Open Science and what their basic requirements are
- be aware of the potential benefits of practicing open science

Open Science definition

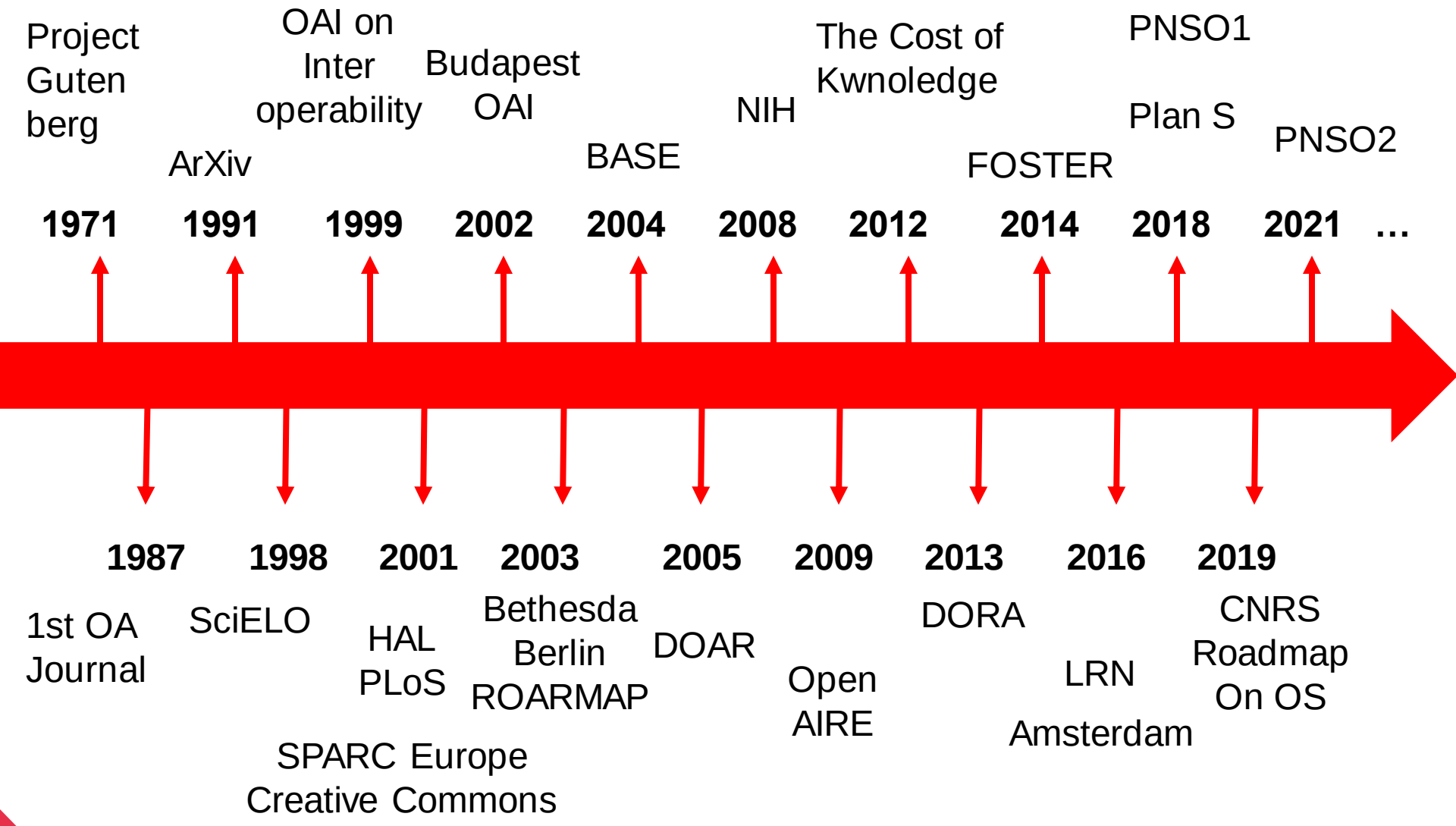
Open Science is the free dissemination of the results, methods and products of scientific research. It is based on the opportunity the digital transformation represents to help develop open access to publications and – as far as possible – to data, source codes and research methods.

https://www.ouvrirlascience.fr/get-started-with-open-science/

Open Science is the practice of science that allows others to 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.

Open Science Definition | FOSTER (fosteropenscience.eu)

OPEN SCIENCE TIMELINE



Amsterdam Call for Action on Open Science

This Call for Action is the main result of the Amsterdam conference on **"Open Science– From Vision to Action"** hosted by the Netherlands' EU presidency on 4 and 5 April 2016.
It formulates two important pan-European goals for 2020 :

1. **Full open access for all scientific publications**
2. **A fundamentally new approach towards optimal reuse of research data**

To reach these goals by 2020 :

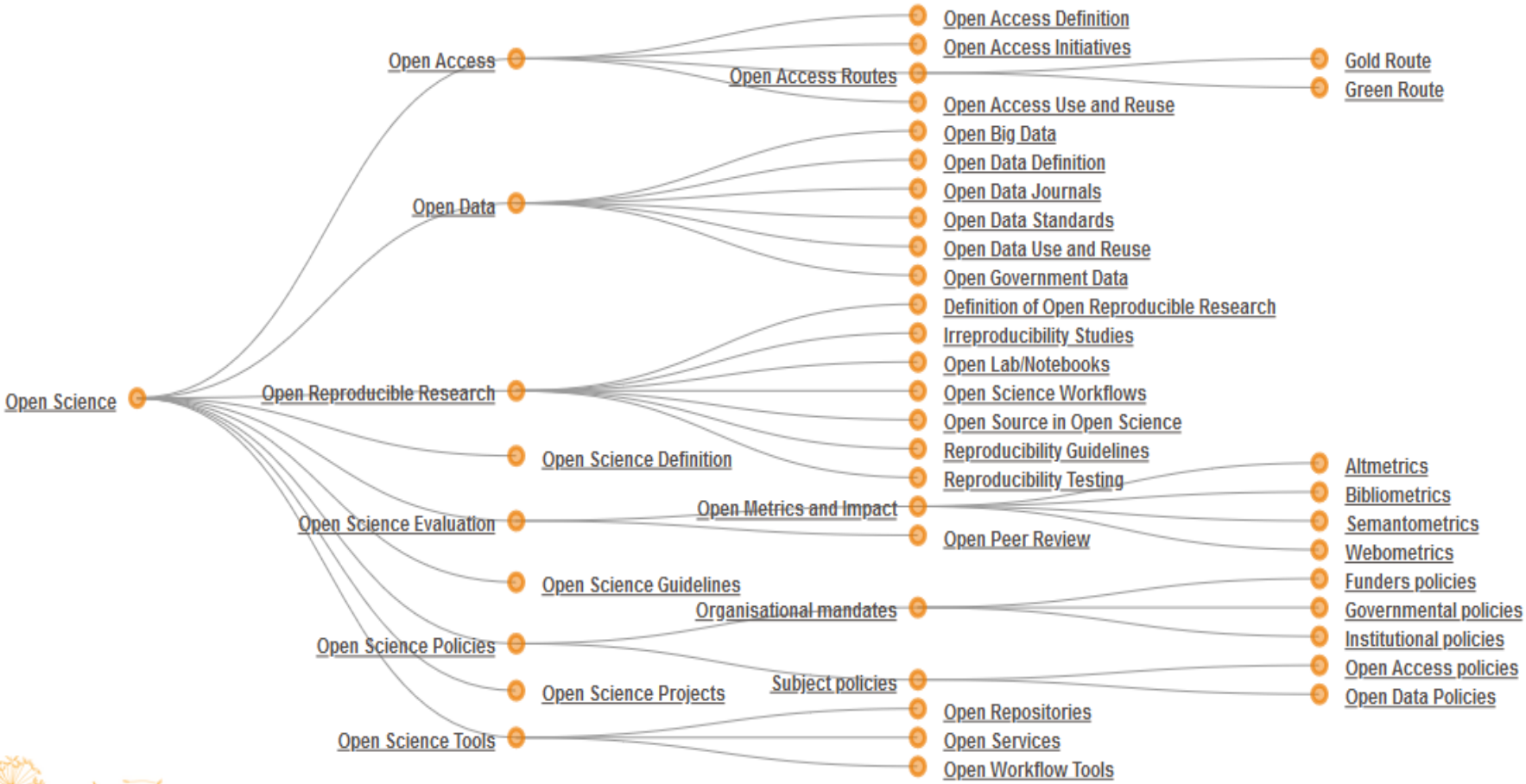
1. **New assessment, reward and evaluation systems**
2. **Alignment of policies and exchange of best practices**

OPEN SCIENCE : a quick view

Components of Open Science

https:///soc.kuleuven.be/mintlab/blog/news/opensciencediscourse

OPEN SCIENCE SCOPE

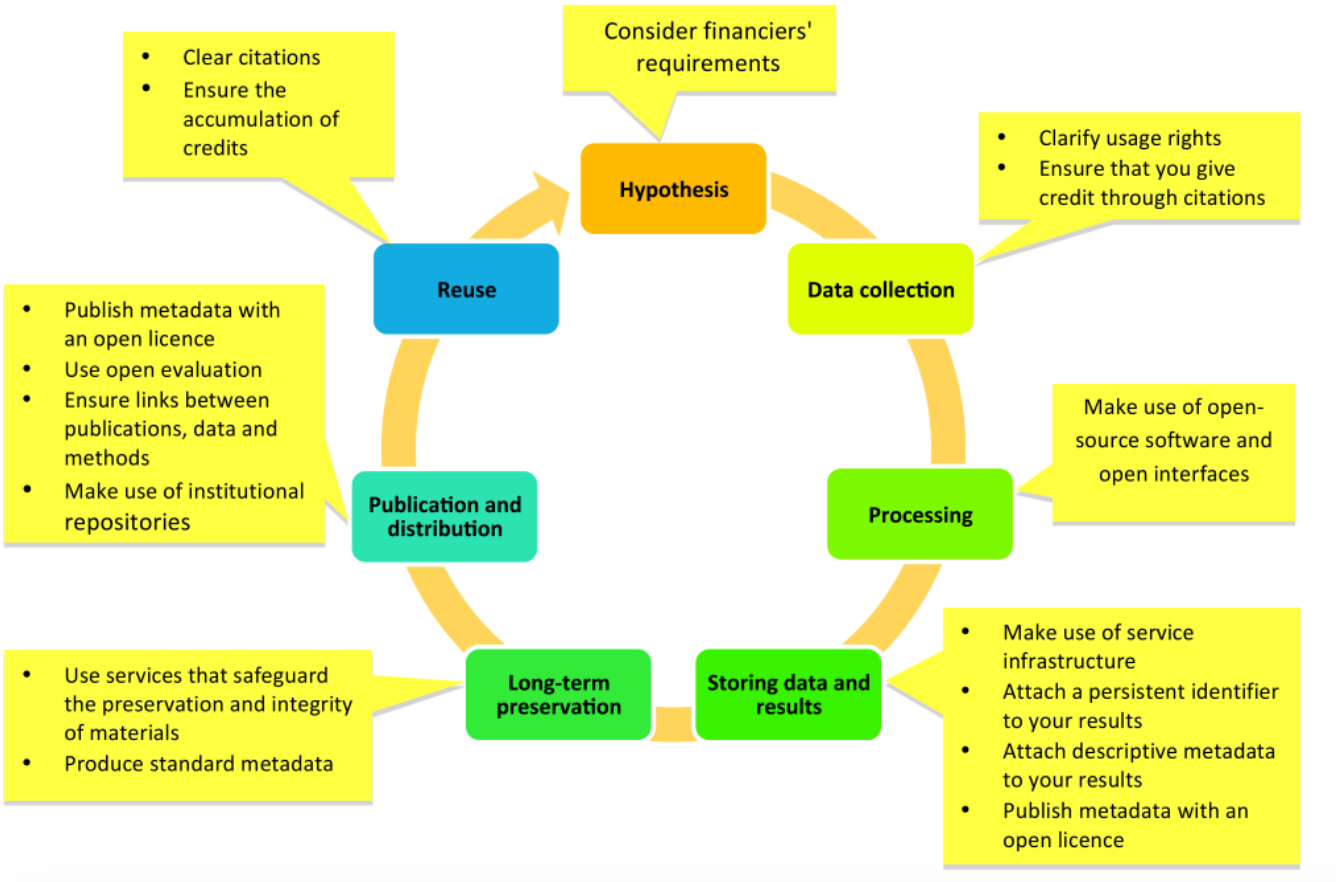


FOSTER

https://www.fosteropenscience.eu/foster-taxonomy/open-science?page=37

Opening the whole research cycle

Open Science is about **extending the principles of openness to the whole research cycle**



Promoting openness at different stages of the research process
https://www.fosteropenscience.eu/content/what-open-science-introduction

OPENING SCIENCE : why ?

Academic issues

- Ease access to scientific and technical information
- Ease collaboration, contribution, circulation of ideas,… among peers
- Warrant Knowledge objectivity

Image CC-BY-NC-SA by Tom Magllery https://flic.kr/p/4FUTKA

Political issues

- Restoring trust (anti-science movements)
- Science at the heart of democracy
  - need for citizen information
  - citizen science : in collecting datas, in opening your results

Economic issues

- Efficiency of public spending
- Knowledge economy
  - sharing generates more growth and cost reduction

PLAN

1. What is Open Science ?

2. Open Access
   - Definitions
   - National & international context

3. Research Data

4. UM and Open Science

Definition : Open Access

Budapest BOAI :

*"…its **free availability on the public internet**, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. … to give authors control over the integrity of their work and the right to be properly acknowledged and cited."*

GREEN OA :
self-archiving on instutional depository

GOLD OA : Open journal With peer-reviewing, free access to publications, **on the publisher's website**, with APC or not

https://www.budapestopenaccessinitiative.org/

Gold Open Access Journal

An open access journal is a scientific journal with an editorial board and review process, which provides access to **published articles free of charge for the reader**. Here are 4 economic models :

Hybrid
Subscription + Article Prossecing Charges (APC) = variable amounts
Absolutely avoid!

Author-pays
Publication costs to authors (APC) = variable amounts. Ex : PLoS
Check predatory editors in DOAJ

Freemium
Access to free html but value-added services reserved for subscribers (pdf, stats, etc.)
Ex: OpenEdition

Diamond
Upstream funding by an institution (learned society, university, library, association, etc.)

Publish in Gold Open Access Journal

DOAJ is an international index of peer-reviewed Open Access Journals selected regarding principles of transparency and their editorial quality

17 420 journals

12 275 without APC

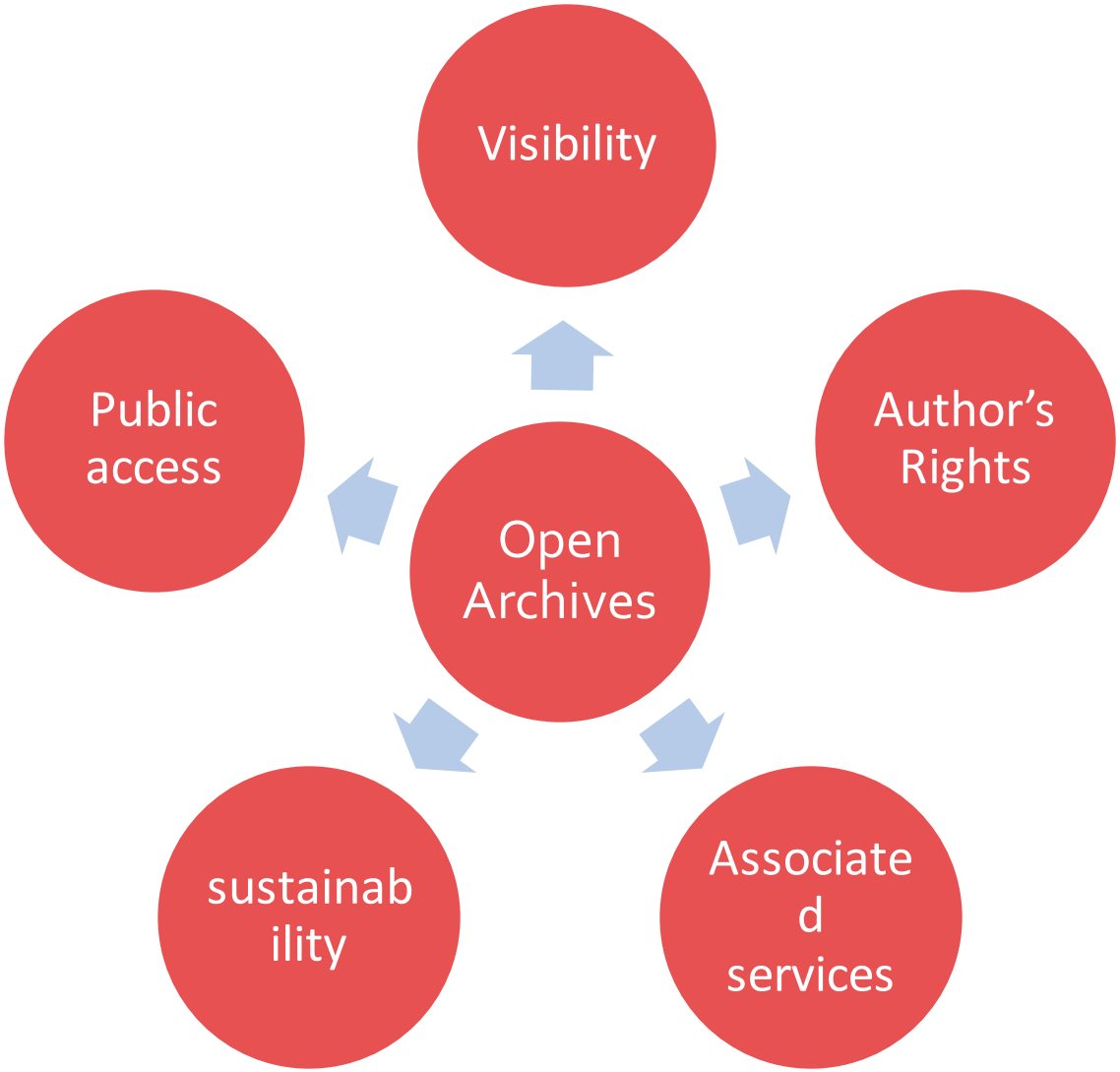
10% have been awarded the DOAJ Seal (7 criteria)

Green Open Access = Open Archives

Open archives are OA repositories where the electronic versions of academic productions published or not (articles, theses, preprints,..) and, in some cases, data obtained from scientific research, can be deposited.

- International : **DART-Europe** - **HAL-Francophonie Afrique et Océan Indien**
- Nationale : HAL France, RCAAP Portugal…
- Institutionnal : Archimer ARThèque
- Scientific subject or aggregating :

Open archives' benefits



PLAN

1. What is Open Science ?

2. Open Access
   - Definitions
   - National & international context

3. Research Data

4. UM and Open Science

French Law for a Digital Republic (2016)

Goals :
Protect authors
Promote the free flow of ideas

https://www.lexplicite.fr/la-loi-republique-numerique-pouvoirs-arcep/

Means :
Introduce a secondary exploitation right for the benefit of authors

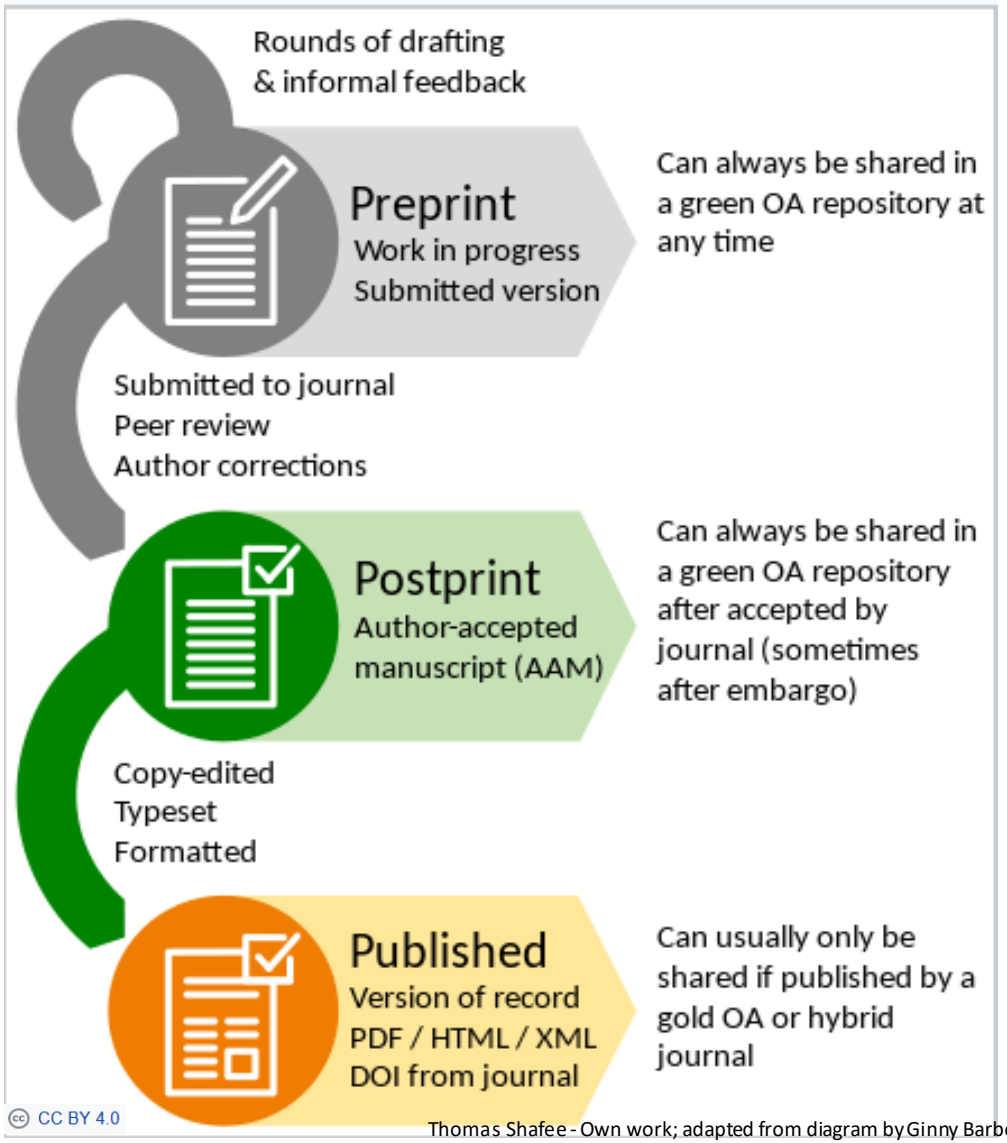
Research publications funded at least at 50% by public funds can be made available free of charge (Green OA) after an embargo period (6 to 12 months), irrespective of academic publishers' policies.

French Law for a Digital Republic (2016)

Which version

for

self-archiving ?



Thomas Shafee - Own work; adapted from diagram by Ginny Barbour

National context

National Plan for Open Science 2018

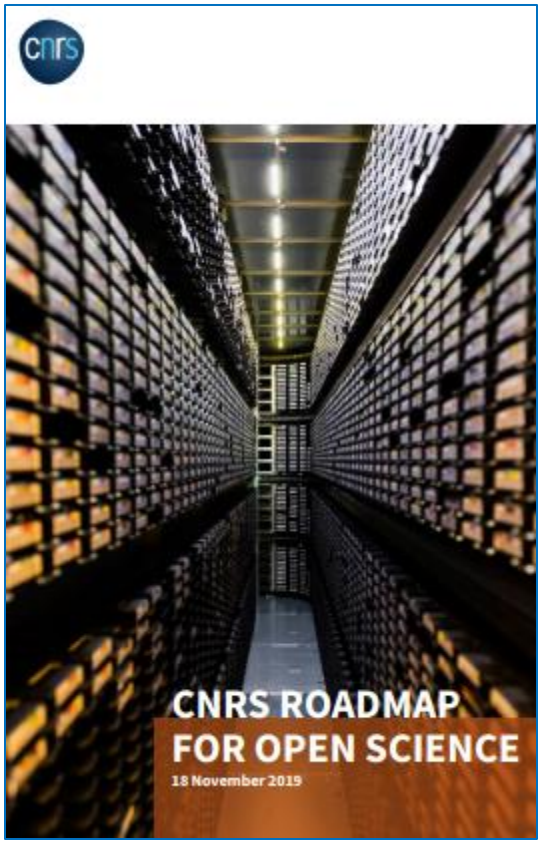
- Generalise open access to publications
  - OA publication obligation for articles and monographs
  - Institutional recognition of open science practices
- Structure and open up research data
  - Obligation to open data from public-funded projects
  - Generalize Data Management Plan (ANR 2019)
- Promote the development of open science practices
  - Generalize "Open Science" skills : doctoral schools labeling)

2d National Plan for Open Science 2021

- Evolve toward Open Science by default

https://www.ouvrirlascience.fr/

CNRS Roadmap Open Science



CNRS Roadmap

1. Publications
2. Research data
3. Text and data mining and analysis
4. Individual evaluation of researchers and Open Science : jso-cnrs-2021
5. Recasting Scientific and Technical Information for Open Science
6. Training and skills
7. International positioning

European context : Plan S by cOAlition S

cOAlition S is a consortium of

national research agencies and funders

from twelve European countries.

The plan requires scientists and researchers who benefit from state-funded research organisations and institutions to publish their work in **open repositories or in journals** that are available to all by 2021

Immediate Open access

=

Attribution license CC BY

Feb. 4th & 5th 2022 : Paris Open Science European Conference

https://osec2022.eu/paris-call/

PLAN

1. What is Open Science ?
2. Open Access
3. Research Data
    1. Definition
    2. Data Management Plan (DMP)
    3. Data repository
4. UM and Open Science

Open Research Data : definition

Research data is any information that has been collected, observed, generated or created to validate original research findings.

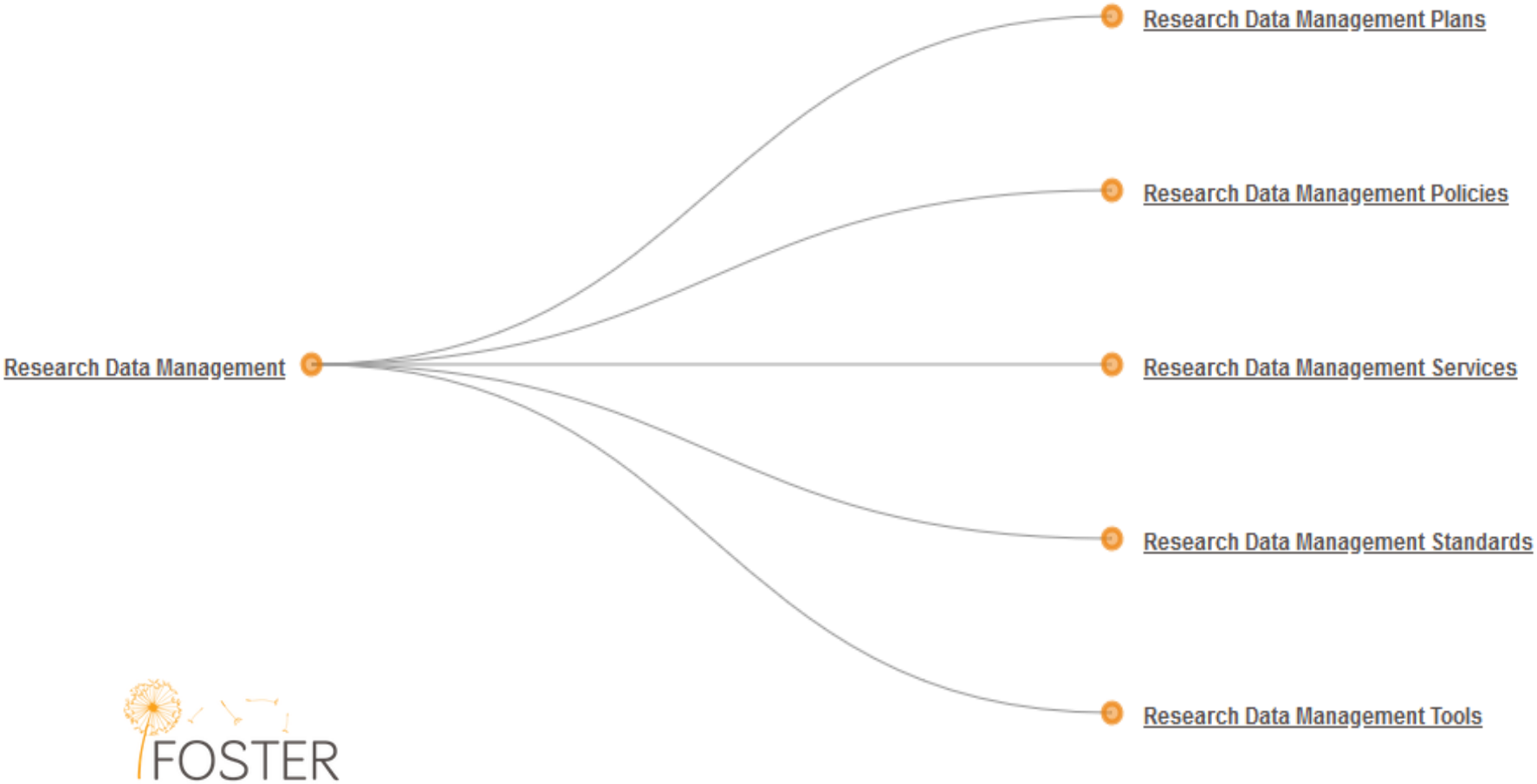
Open research data refers to the data underpinning scientific research results that has no restrictions on its access, enabling anyone to access it.

(Facts and Figures for open research data)

They can be different in :

- **media:** lab notebooks, electronic documents, software, computer programs…
- **types:** archives, audio, video, databases, source codes, geospatial data, images, photographs, digitizations, scans, qualitative, quantitative, statistics…
- **natures:** raw, experimental, produced or collected, from observation or simulation...

Research Data Management



FOSTER

The research data lifecycle

ETH Data lifecycle

What if… you lost all your data?

As much as 80% of research data is currently lost within 20 years

A loss for:

- Authors and his/her team, organizations, institutions
- global research

Reasons:

- Destruction of supports
- Version problem,
- hardware or software obsolescence
- Indefinite storage location

Consequences :

- Waste of time, money (public funds, etc.)
- No verification of results possible
- No comparison of results in time or space
- No (different) reuses

[The Availability of Research Data Declines Rapidly with Article Age]

What is at stake in sharing research data

- **Heritage** issues

- **Economic** issues
  - Economic value of the data
  - Reuse of data and exploitation of research results
  - R&D : accelerate innovation (faster to market = faster growth);

- **Scientific** issues
  - Foster collaboration and avoid duplication of effort (greater efficiency)
  - Build on previous research results (improved quality of results)
  - Improves the integrity of the scientific and scholarly record
  - More visibility and citations for the scientist and his/her team

- **Social** issues
  - involve citizens and society (improved transparency of the scientific process).

Hrynaszkiewicz, I., Harney, J. and Cadwallader, L., 2021. A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Science Journal*, 20(1), p.31. DOI: http://doi.org/10.5334/dsj-2021-031

What are the Findable Accessible Interoperable Reusable data principles ?

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE

Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

Association of European Research Libraries

PLAN

1. What is Open Science ?
2. Open Access
3. Research Data
    1. Definition
    2. Data Management Plan (DMP)
    3. Data repository
4. UM and Open Science

Data Management Plan (DMP)

A DMP is a document that describes

- which data you expect to acquire or generate during the course of a research project
- who is responsible and who collaborates
- how you will manage, describe, analyze, and store those data
- what mechanisms you will use to share
- where preserve your data

It helps you **formalize** the process and provide **you and your collaborators** with a **trustworthy** document all along the **research progress**

Data management is best addressed in **the early stages** of a research project, but it is **never too late** to develop a DMP

DMP : you are not alone

Different actors and skills

- Researcher : production process and scientific field
- Data librarian or ingeneer : metadata, warehouses, archiving
- IT : data security
- Legal : IP of the data
- Project engineer : agreement, eligibility of costs

DMP and funding

Funding agencies and sponsors, public and private, require a DMP **covering the lifespan of a research project** as part of grant proposals. These plans typically state what data will be created and how, specify **who is responsible** for those data, and outline the plans for **preservation and sharing**, noting what is **appropriate to share** given the nature of the data and any **restriction**s that may need to be applied

Each agency or directorate creates its own set of policies for data management and also according to the disciplinary calls for project funding : **Horizon Europe, ANR**

https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1

European Commission, Directorate-General for Research and Innovation, *The economic rationale for public R&I funding and its impact*, Publications Office, 2017, https://data.europa.eu/doi/10.2777/047015

DMP in practice

Many DMP templates by funding agencies ANR / institutions UM Most follow a six-levelled-structure with the same items :

- Administrative information
- Data description: identify, organize, document the types of datasets collected and generated in the project
- Documentation, metadata, standards
- Data storage during the project
- Sharing data in a repository
- Permanent archiving
- Data Security
- Legal and ethical issues: standards, charters, ethics committee, intellectual property, copyright
- Responsibilities
- Costs

2 collaborative and free tools

INIST-CNRS OpenAire

W.K. Michener
Ten Simple Rules for Creating a Good Data Management Plan
https://doi.org/10.1371/journal.pcbi.1004525

PLAN

1. What is Open Science ?
2. Open Access
3. Research Data
    1. Definition
    2. Data Management Plan (DMP)
    3. Data repository
4. UM and Open Science

Collect data from… Data repositories

According to the Registry of Research Data Repositories
a Data Repository is :
« A subtype of a **sustainable information infrastructure** which provides **long-term storage and access** to research data and source codes. »

It assignes a **DOI** to each uploaded object and provide a web page that tells what it is, **how to cite** it and how many times other researchers have cited or downloaded that object

A Data Repository can be
- generalist or disciplinary
- public or proprietary

Data repositories provide a **search engine** and **filters** to select the data sets you would like to use in your research

Data storing and sharing

Why does preservation matter ?

- Technological obsolescence
- Disappearance of formats
- No reliable supporting documentation or metadata

Determine
the level of access

How to share my data sets ?

As studies are completed and their data become available some information will be posted per protocol, such as

- Study protocol
- Reference to study publication of primary outcome
- Data sets access and formats
- Annotated case report forms
- Define file (also known as Data Dictionary)
- Study-specific de-identification notes
- …

Types of data repositories

Disciplinary

VirHosNet : **virus/host molecular interactions data**
SEANOE: French national repository for marine research data
PANGAEA : international repository for earth system science
TCIA : The Cancer Imaging Archive part of the US NCI

Mutidisciplinary

ZENODO: both publications and data
Figshare : All file formats can be published
Dryad : data underlying medical literature
***Recherche Data Gouv*: national research data platform**
Github : the largest code host in the world

Institutional

Edinburgh DataShare
Merritt
Enlighten University of Glasgow
Dataverse CIRAD

Mendeley Data (Elsevier)
Gigascience/GigaDB (OUP and BGI)

Trusted repositories

WHY ?

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

WHO ?

CoreTrustSeal is an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures. France : **Strasbourg Astronomical Data Center** and **CINES**

HOW ?

The organization has determined requirements to make sure a repository can be trusted in the long term, among which :

- Access right information
- Curation
- Digital preservation
- Integrity
- Migration

Data repositories major repertories

Multidisciplinary

Re3data (German Research Foundation DataCite)

OAD (USC Libraries)

Spécific to a research area

FAIRsharing (biomedical & life science)

NIH Data Sharing Repositories (Biomedical)

CAT Opidor (French repositories))

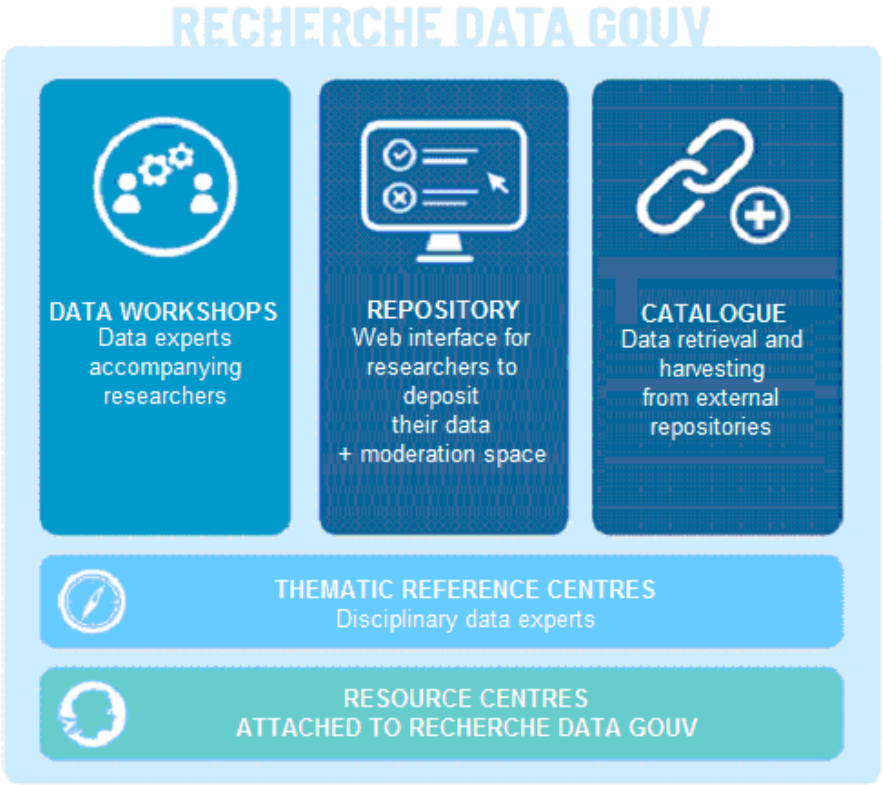
Generic

OpenDOAR

ROAR

The french national repository : *Research Data Gouv*

The national federated research data platform will be available in the first quarter of 2022



https://www.ouvrirlascience.fr/recherche-data-gouv-the-federated-national-research-data-platform/

PLAN

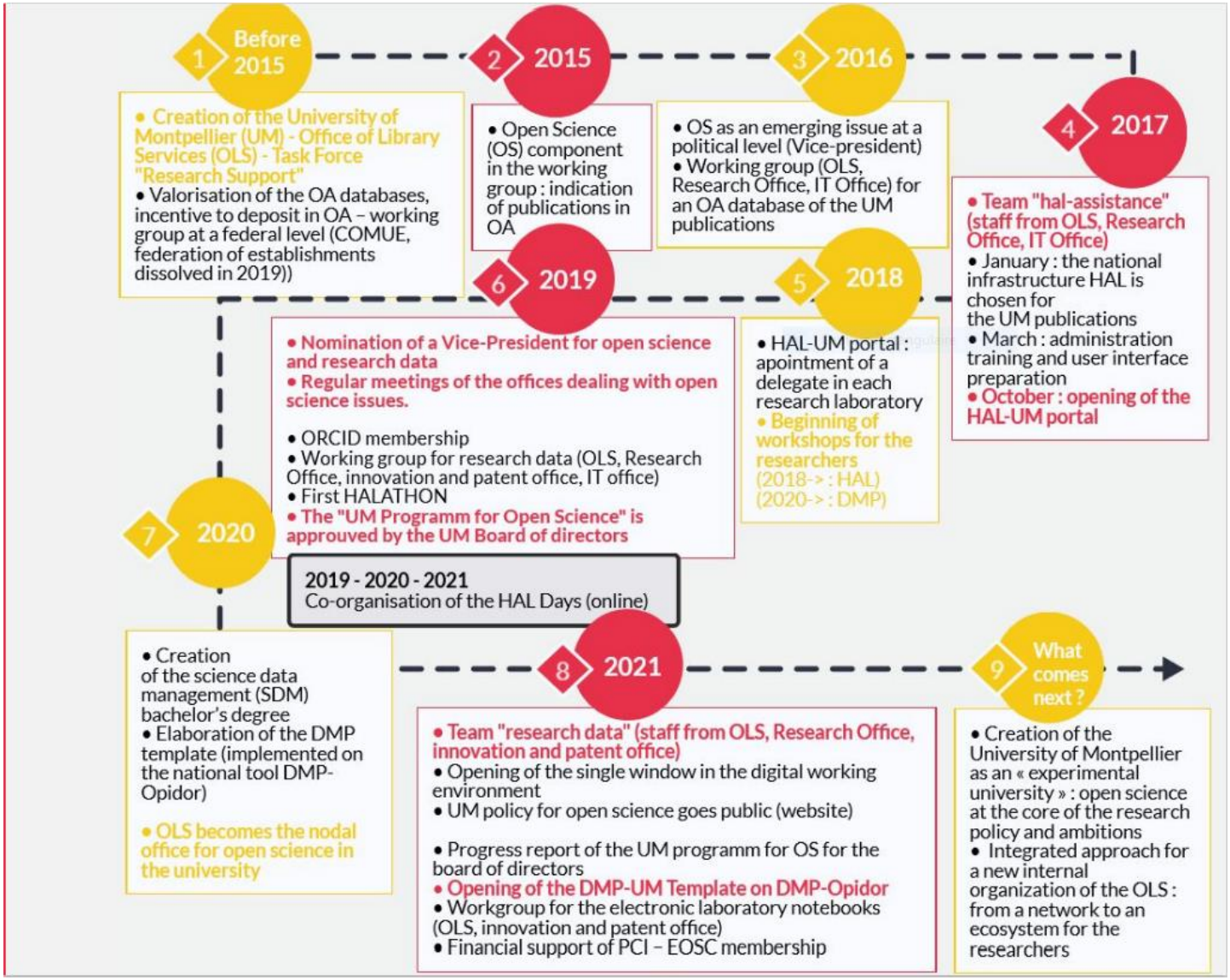
1. What is Open Science ?

2. Open Access

3. Research Data
    1. Definition
    2. Data Management Plan (DMP)
    3. Data repository

4. UM and Open Science

UM and Open Science



The UM programm for Open Science

Nomination of a **Vice-President for Open Science and research data** (2019), Anne Laurent

Aims of the UM programm for Open Science : have a global open science approach and promote responsible management of research data

The **UM programm for Open Science** has been approved in 2020 by the UM boards of directors and its progresses are regularly evaluated :

- Affirm the institution's commitment (signature of the Berlin's declaration and of the Jussieu's Call)
- Establish the use of HAL in all communities and increase the full-text repository in the HAL-UM portal https://hal.umontpellier.fr/
- Promote responsible data management by the assistance on how to fill a Data Management Plan and with a UM DMP model, storage infrastructures, helpdesk
-Provide data and results protection tools and procedures : electronic laboratory netbook, sensitive data storage

Disseminating the publications of the UM's researchers : the HAL-UM portal

https://hal.umontpellier.fr/

A team to help any researcher or laboratory how to deposit or disseminate their scientific publications :
HAL-Assistance : hal-assistance@umontpellier.fr

A assistance to manage research data

A team to help researchers to fulfill their DMP
Données de la recherche : donnees-recherche@umontpellier.fr

Experts at the Institut of Data Science of Montpellier
https://isdm.umontpellier.fr/

Scientific Data Management bachelor's degree
https://sdm.edu.umontpellier.fr/

Storage infrastructure
Meso@LR https://meso-lr.umontpellier.fr/
Meso2-Data
(Meso3-HDS)

Thank you for listening !

Any questions ?

https://bibliotheques.edu.umontpellier.fr/science-ouverte/

Contacts : odile.hennaut@umontpellier.fr and laure.lefrancois@umontpellier.fr