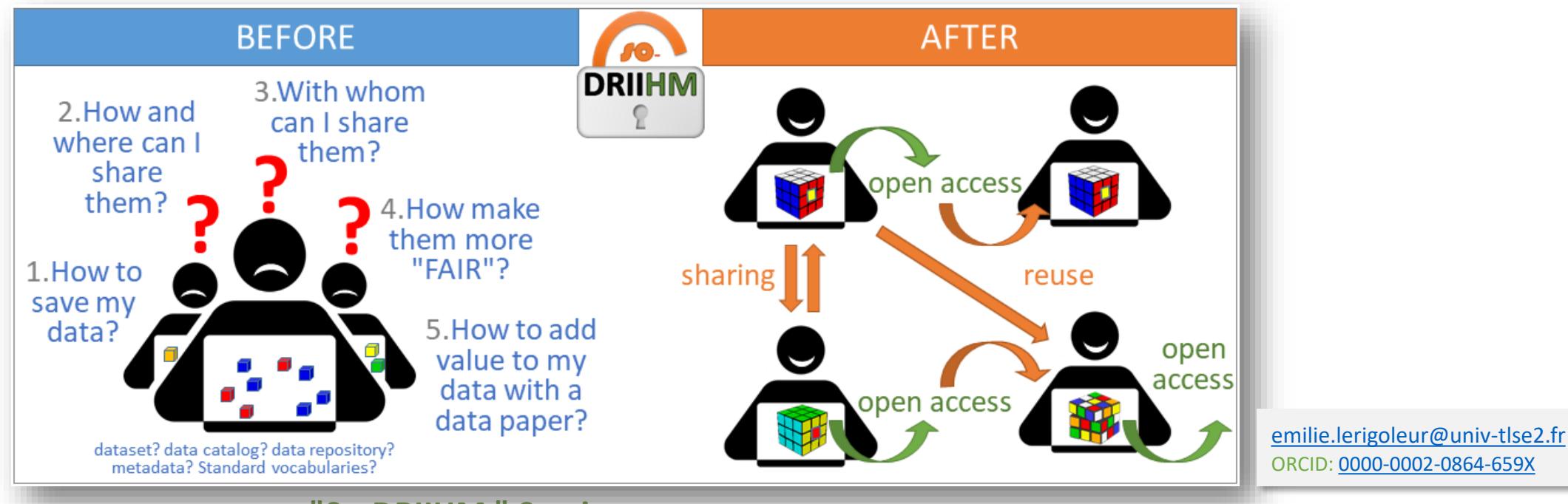


Vision and first actions of the SO-DRIIHM open science project

Promote Open Science within the DRIIHM Interdisciplinary Research Facility on Human-Environment Interactions: Co-design of an e-infrastructure implementing the FAIR Principles

Emilie Lerigoleur, Ghita Amal, Fanny Arnaud, Pascal Dayre, Armelle Decaulne, Priscilla Duboz, Samuel Dunesme, Arnaud Jean-Charles, Iwan Le Berre, Mathieu Massaviol, Kristell Michel, Bertrand Morandi, Corinne Pardo, Jean-Claude Raynal, Grzegorz Skupinski, Marie-Laure Trémélo



1. Open Science is a growing part of the research world

“Open science refers to the unhindered dissemination of results, methods and products from scientific research”
(PNSO2, 2021)



Around the world, researchers are increasingly aware of the value and importance of Open Science

Many bottom-up, stakeholder-driven and self-governed initiatives act for Open Science



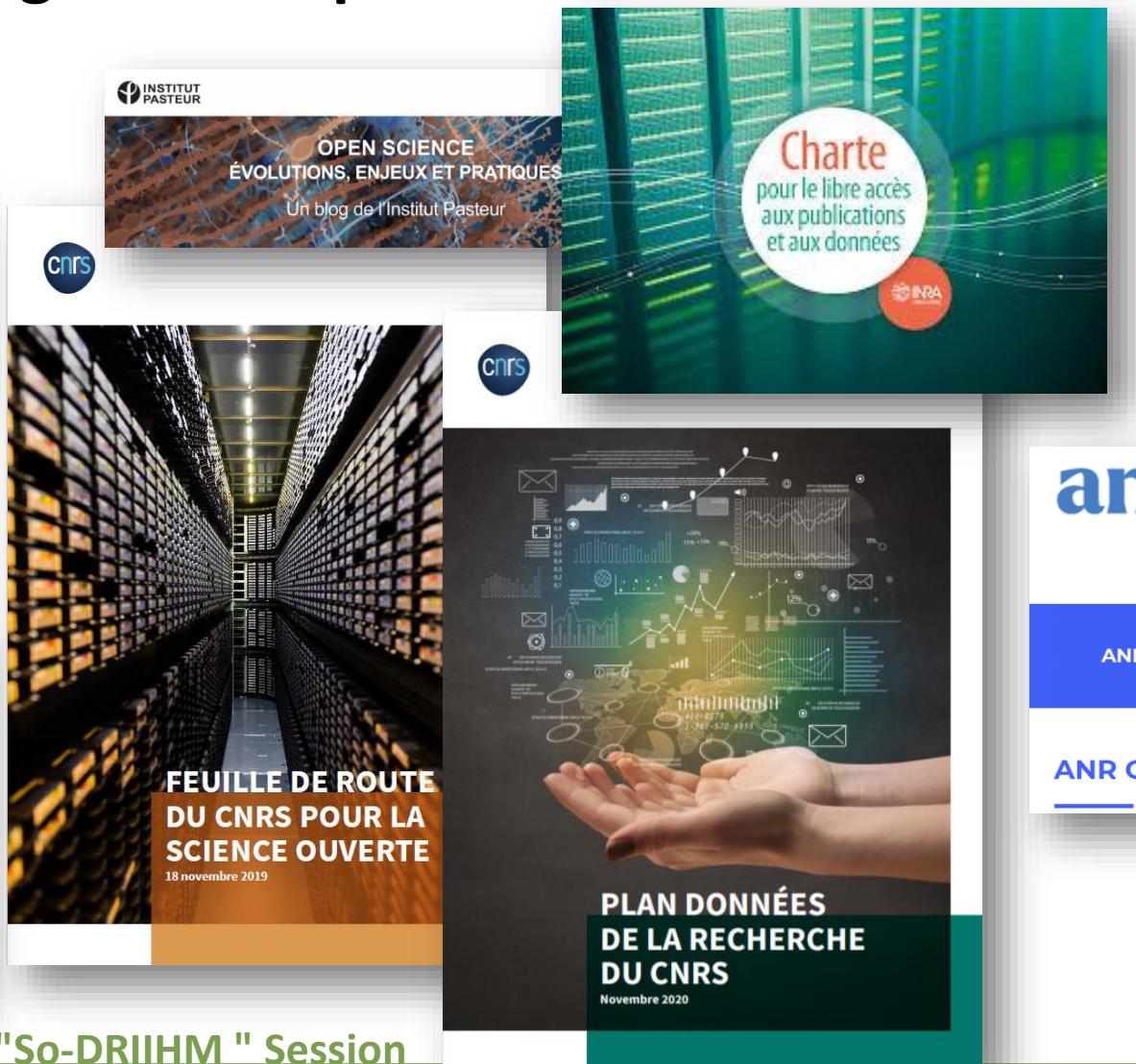
Ex: The European Open Science Cloud ([EOSC](#)) is seeking to make it much easier for researchers to pursue fully open science by federating the myriad of research infrastructures, tools and datasets across the EU



<https://www.rd-alliance.org/data-together-fostering-cooperation-among-open-science-platforms-march-2021>

"So-DRIIHM " Session

2. Universities, public research institutions and public research funding agencies implement an open science policy



RESEARCH PROJECTS' ACTORS

anr agence nationale de la recherche
AU SERVICE DE LA SCIENCE

ANR's role in research Calls for proposals

ANR Open Science policy

3. The 2nd National Plan for Open Science (PNSO2) generalises open science in France

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
*Liberté
égalité
fraternité*

Deuxième Plan national pour la science ouverte



4 themes:

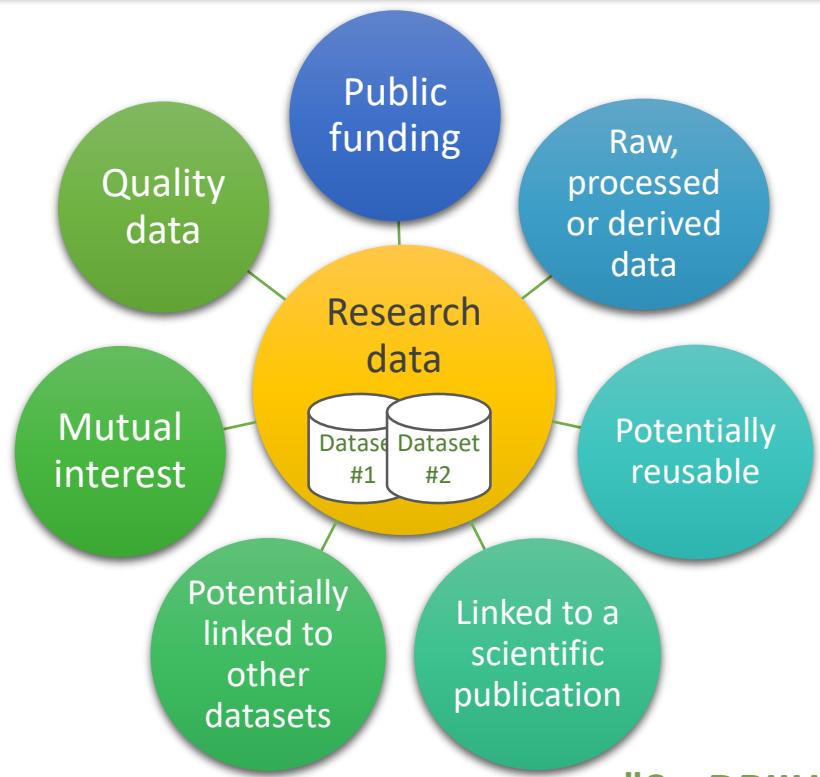
- 1- Generalising open access to publications
- 2- Structuring, sharing and opening up research data
- 3- Opening up and promoting source code produced by research
- 4- Transforming practices to make Open Science the default principle

"So-DRIIHM " Session

4. What research data are we talking about?

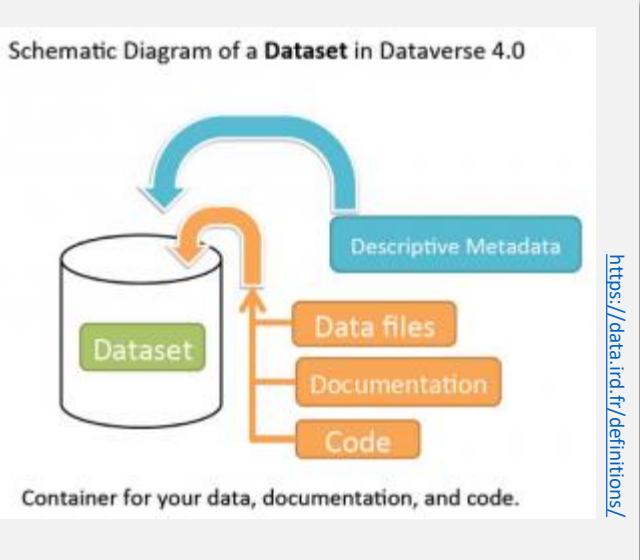
“**research data** are defined as **factual records (numerical scores, textual records, images and sounds)** used as primary sources for scientific research, and that are commonly accepted in the scientific community as **necessary to validate research findings**.

A **research data set** constitutes a systematic, partial representation of the subject being investigated.” ([OECD, 2007](#))



Each **dataset** (= data set) contains descriptive metadata and data files

→ potentially including documentation and code that accompany the data



"So-DRIIHM " Session

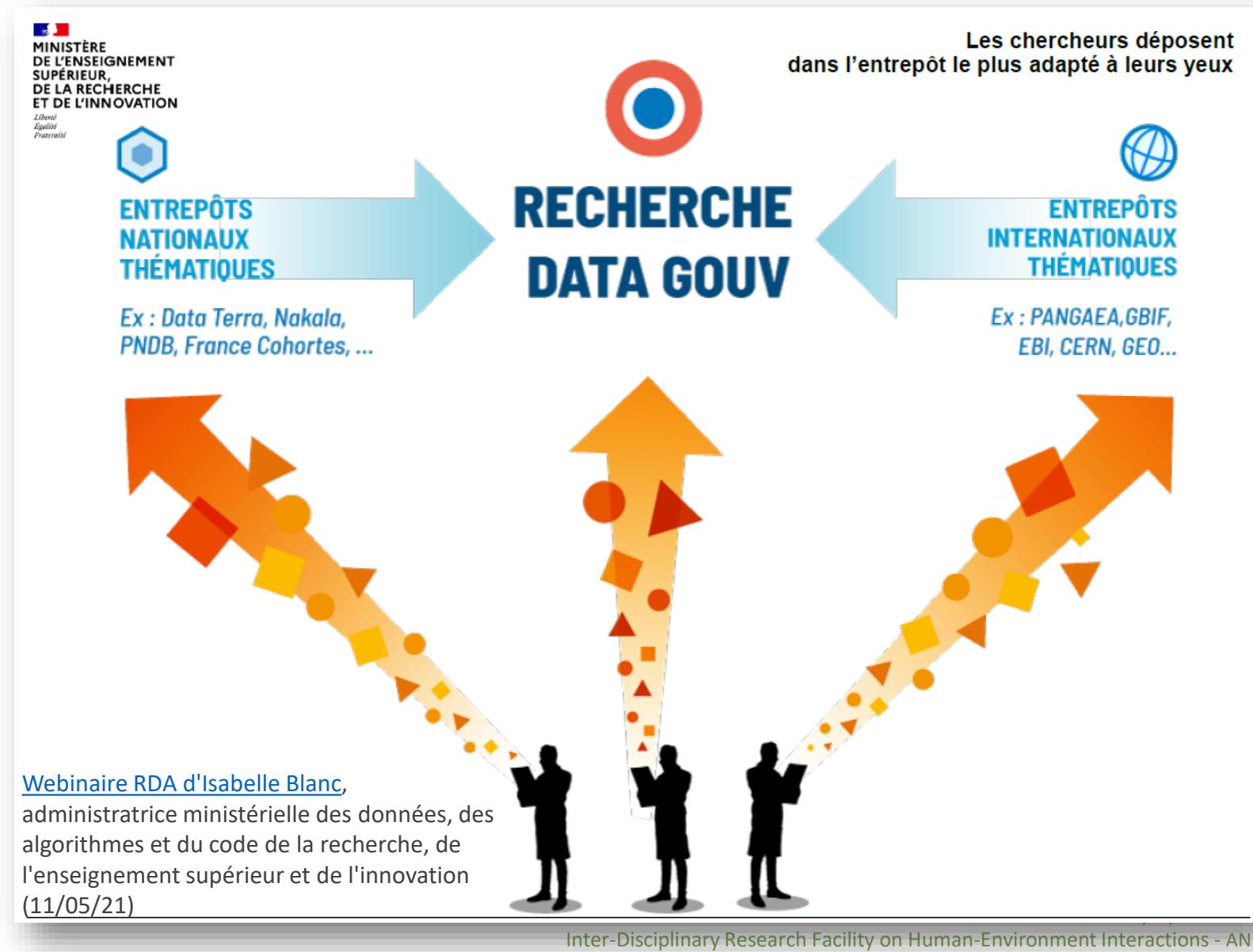
5. Recherche Data Gouv will provide a multidisciplinary data repository

→ to complement the national and European infrastructures already used by some scientific disciplines ([PNSO2, 2021](#))

= Measure n°5 of PNSO2

Data catalogue
+ Data repository
+ Support

recherche.data.gouv.fr



6. Data should be “as open as possible and as closed as necessary”

“open” in order to foster the reusability and to accelerate research, but at the same time they should be “closed” to safeguard the privacy of the subjects ([Landi et al., 2020](#))

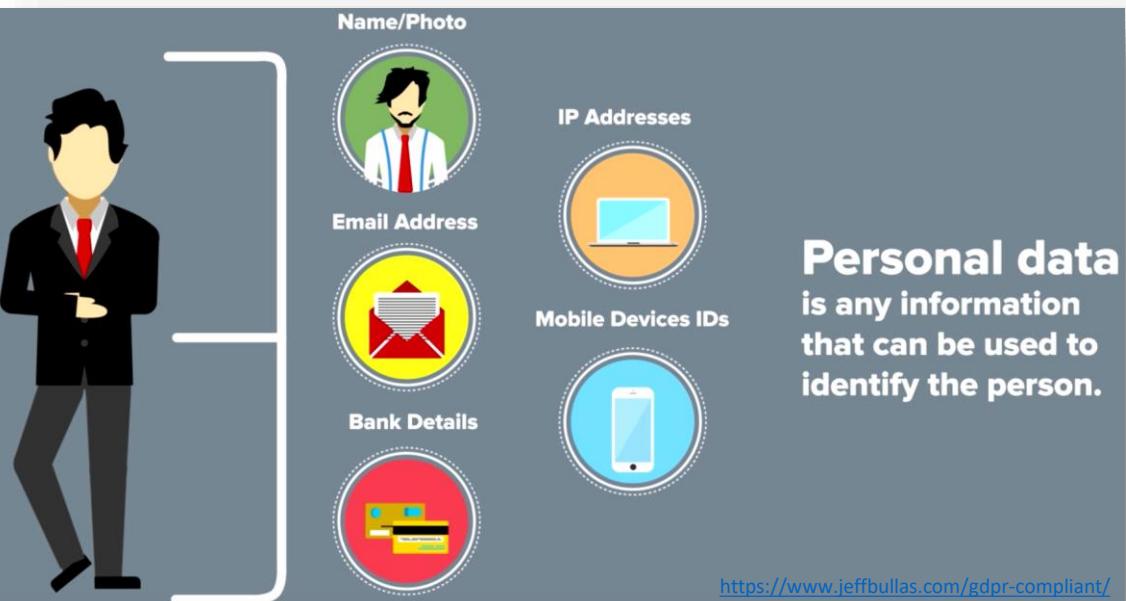
Ex: “exceptions can be made for professional confidentiality, industrial and trade secrets, personal data and copyright-protected content” ([PNSO2, 2021](#))

For European countries:

→ Personal data have been endorsed by the European **General Data Protection Regulation** (GDPR), Reg (EU) 2016/679



“So-DRIIHM ” Session



7. Make data as “**FAIR**” as possible to promote the reuse of them

“data produced by French public research be progressively structured to conform to the FAIR principles, be safely preserved and, wherever possible, open to all” ([PNSO2, 2021](#))

“**FAIR is not equal to Open**: The ‘A’ in FAIR stands for ‘Accessible under well defined conditions’.”

“**Partly FAIR may be fair enough**”
→ many levels of FAIRness
([Mons et al., 2017](#))

Findable

Metadata and data should be findable for both humans and computers

F**A**

Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

I**R**

Accessible

Once found, users need to know how the data can be accessed

Reusable

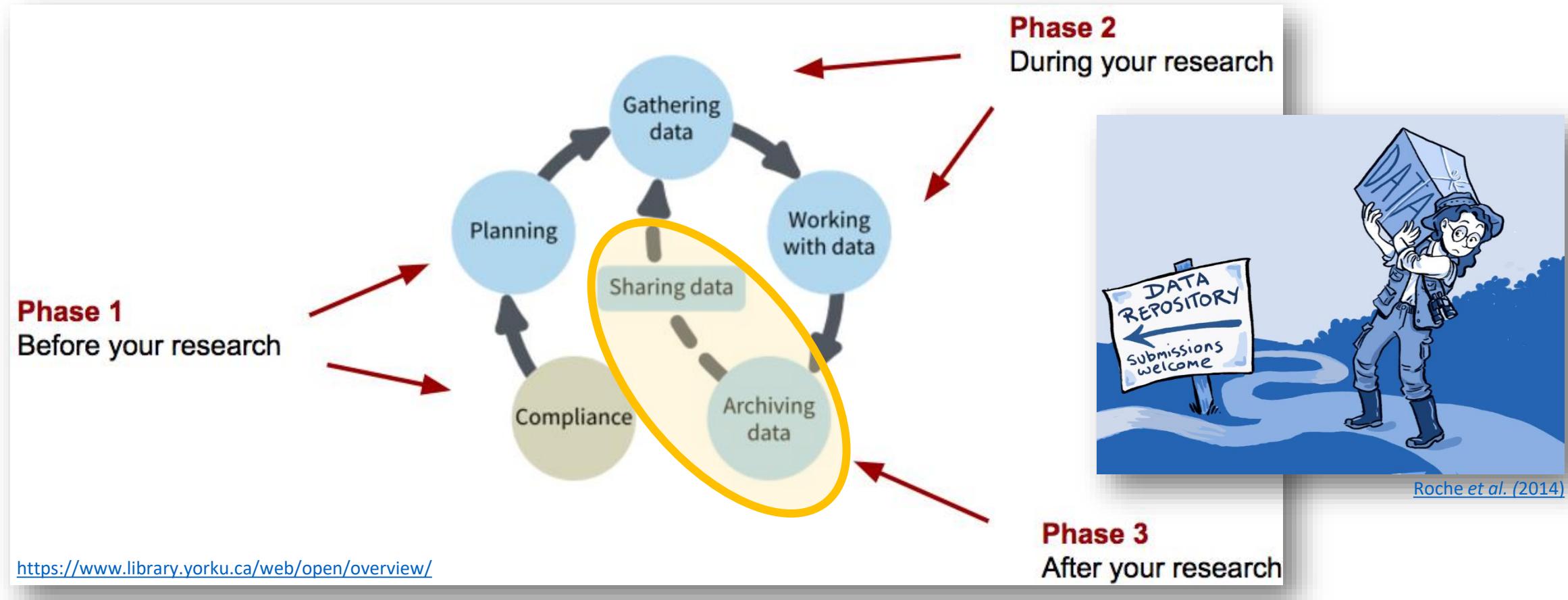
The goal of **FAIR** is to optimise data reuse via comprehensive well-described metadata

<https://www.scibite.com/solutions/enterprise-fair-data-mdm/>

Reference article for FAIR principles:

[Wilkinson et al., 2016](#)

8. Research data preservation, publication and sharing are important but hard to carry out without help



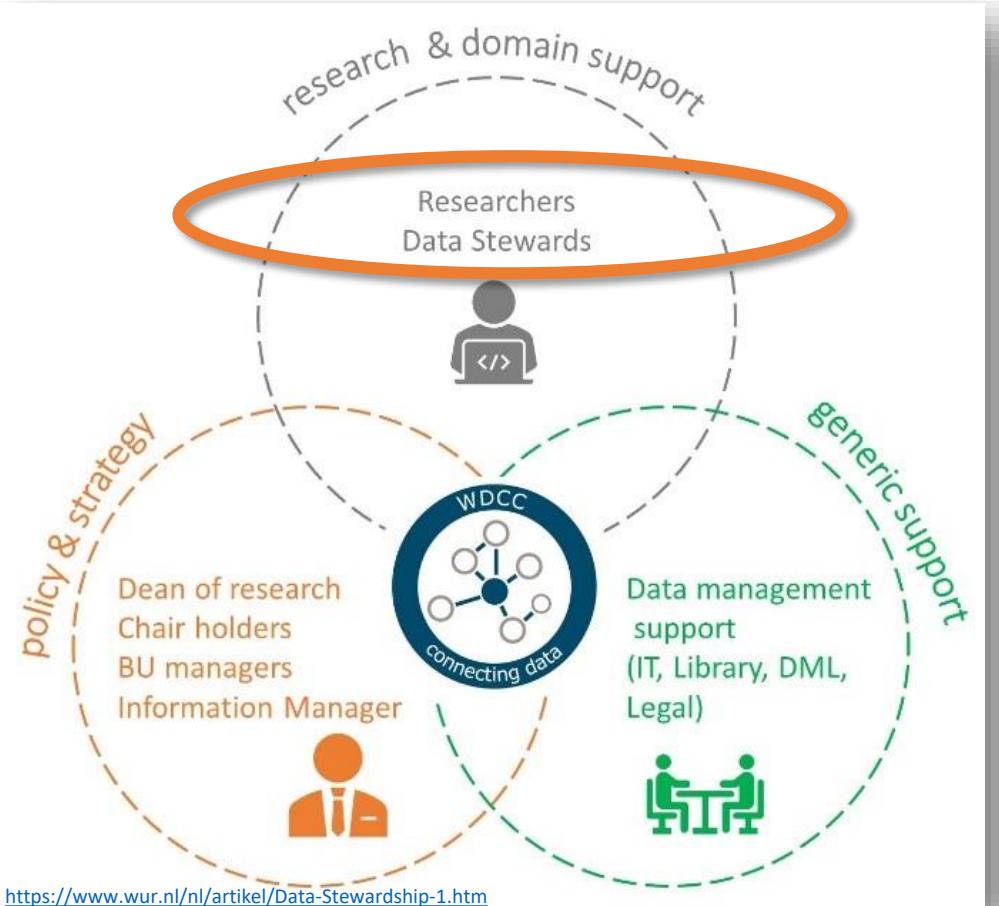
"So-DRIIHM" Session

9. New jobs around data management are emerging but there is still little recruitment

data steward, data champion, data manager, data curator, data librarian, data scientist, data engineer...



Data Stewards: generalists with research background and excellent communication skills



10. Open science... in summary

1. Open Science is a growing part of the research world
2. Universities, public research institutions and public research funding agencies implement an open science policy
3. The 2nd National Plan for Open Science (PNSO2) generalises open science in France
4. Research data (from public funding) = factual records that are necessary to validate research findings
5. Recherche Data Gouv will provide a multidisciplinary data repository
6. Data should be “as open as possible and as closed as necessary”
7. Make data as “FAIR” as possible to promote the reuse of them
8. Research data preservation, publication and sharing are important but hard to carry out without help
9. New jobs around data management are emerging but there is still little recruitment

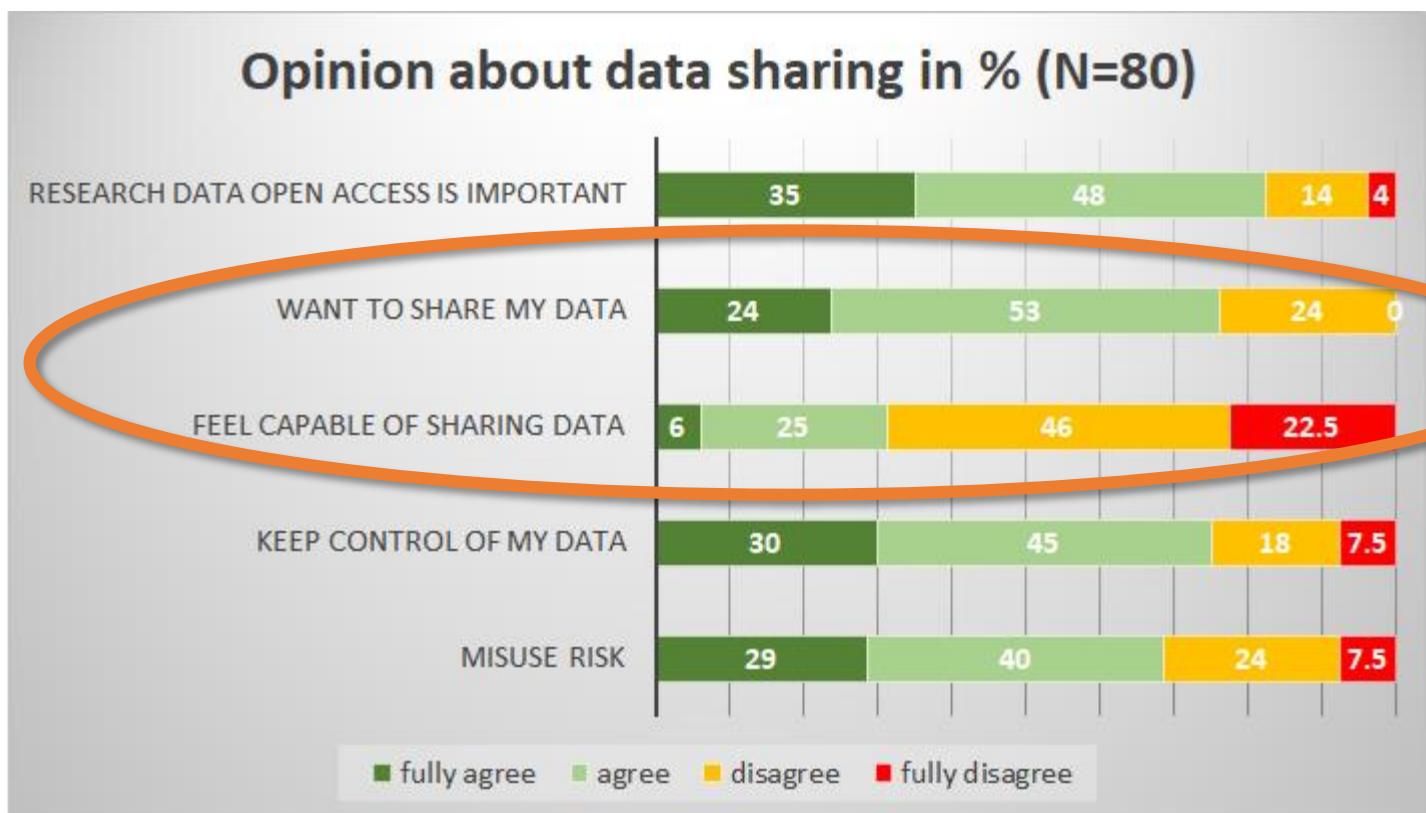
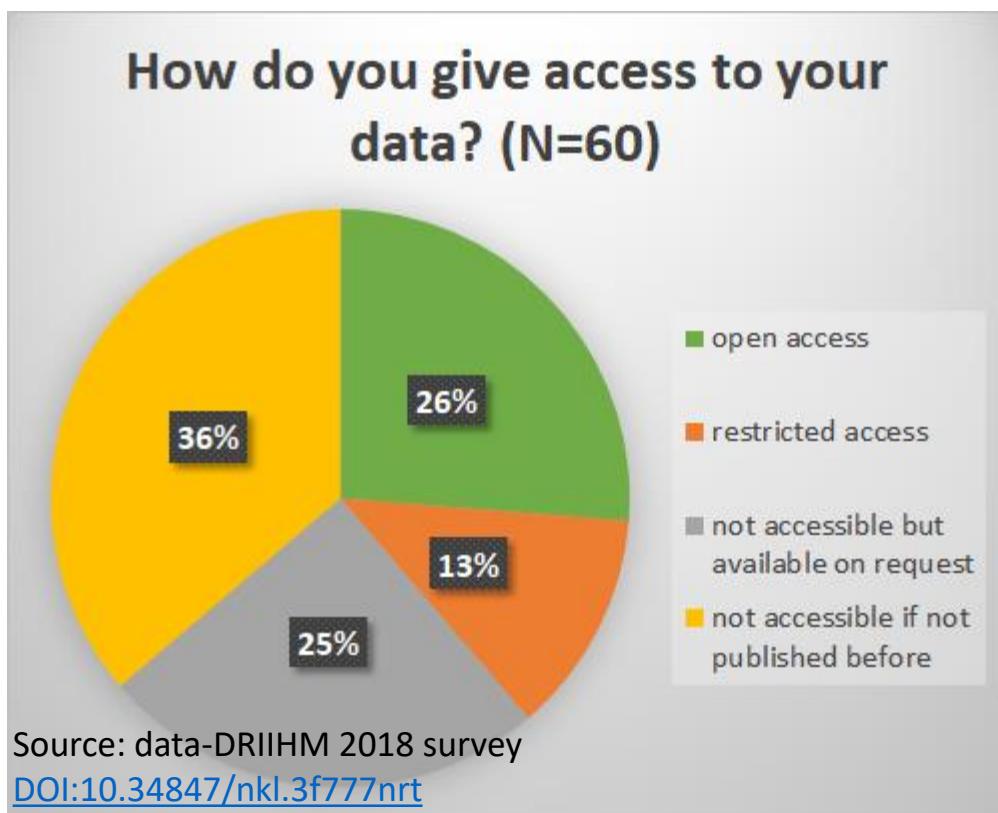
"So-DRIIHM " Session

1. What about open science and data sharing within the DRIIHM?

- Scattered and poorly accessible data and scientific publications
- Risk of loosing valuable observation data
- Little data accessible via geocatalogues and not very user-friendly tools



A step towards sharing and opening up OHM data, but need for support and consideration of potential associated risks



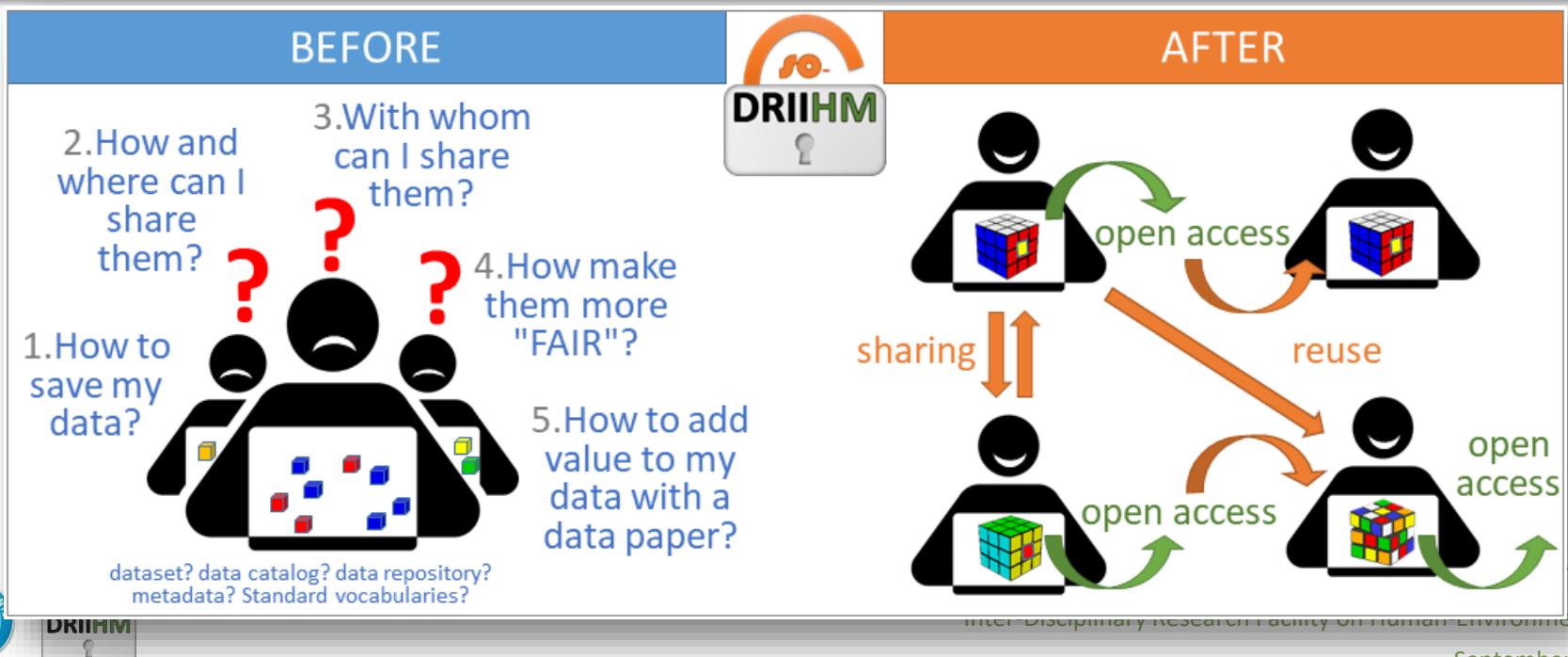
2. The ANR SO-DRIIHM project encourages the sharing and opening of data within the OHMs, between them and beyond

1. Promote Open Science and encourage Open Science practices, showing the challenges and benefits of structuring, accessing, sharing and opening research data

- ✓ ANR Open science Flash 2019
- ✓ 8 partners
- ✓ 5 OHM + 1 OHMi
- ✓ 98 k€
- ✓ Oct. 2020 - March 2023

2. Enhance the DRIIHM existing infrastructure to a more user-friendly and interoperable e-infrastructure with new functionalities that will be prioritized according to users' expectations and motivations

We will help you answer these questions



3. The Data-DRIIHM team with IRIT and Makina corpus at the helm

A. Jean-Charles,
M. Massaviol,
C. Pardo, JC. Raynal

DRIIHM
secretariat and
OHM Bassin minier
de Provence

ML. Trémélo

OHM Littoral
méditerranéen

P. Duboz, *B. Cissé*

Ergonomists and
web development
specialists

OHMi
Téssékéré
(Senegal)

"So-DRIIHM " Session

1. UMR
GEODE
(leader)

2. FR3098
ECCOREV

3. UMR
5600
EVS

4. UMR
7300
ESPACE

5. UMR
6554
LETG

6. UMI
3189
ESS



OHM Pyrénées
Haut Vicdessos

7. UMR
5505
IRIT

OHM
Vallée du
Rhône

+ all
volun-
teers

OHM Littoral
Caraïbe

E. Lerigoleur

P. Dayre

F. Arnaud, *S. Dunesme*,
K. Michel, B. Morandi

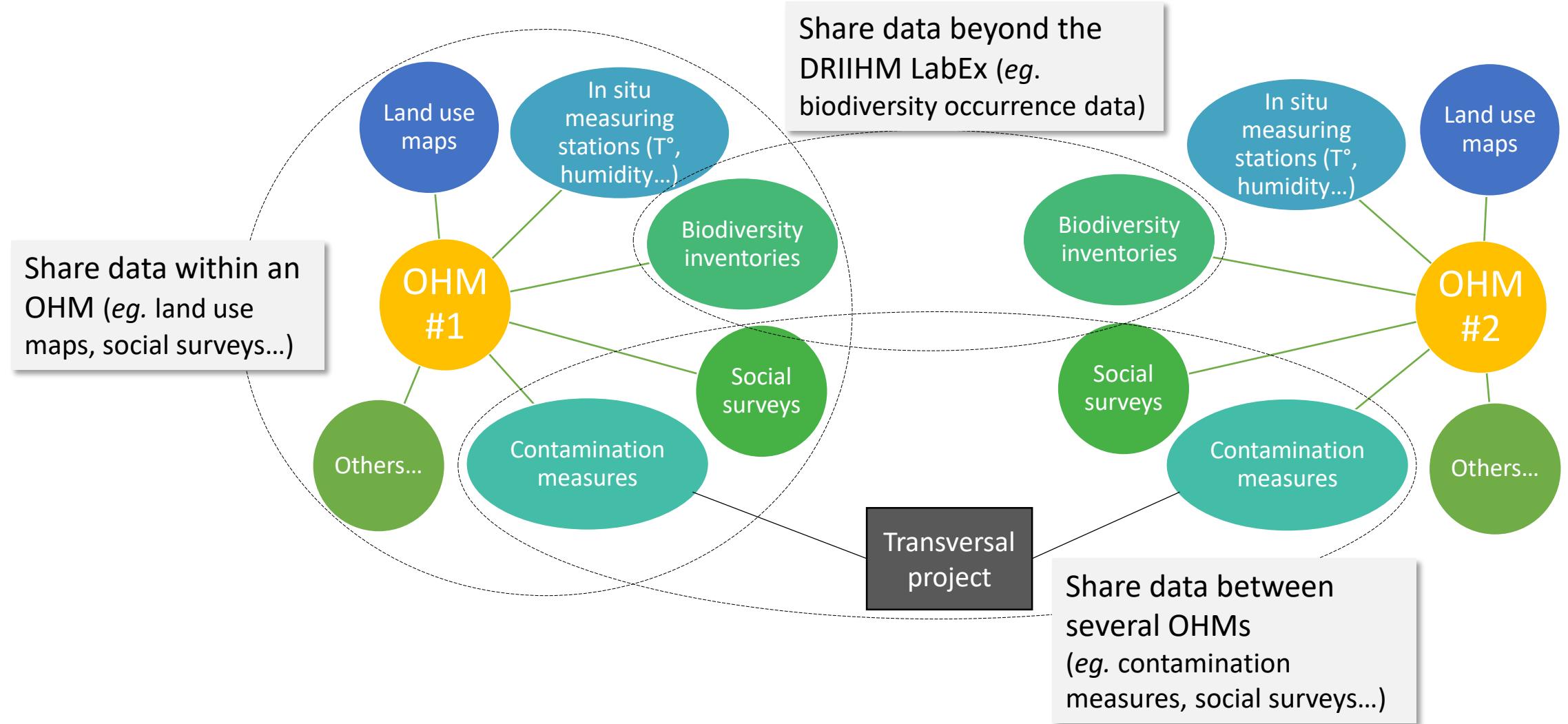
G. Skupinski (OHM Fessenheim)
A. Decaulne (OHMi Nunavik)

I. Le Berre

+ All DRIIHM volunteers among researchers, engineers and
OHM directors

*Italic: New
participants*

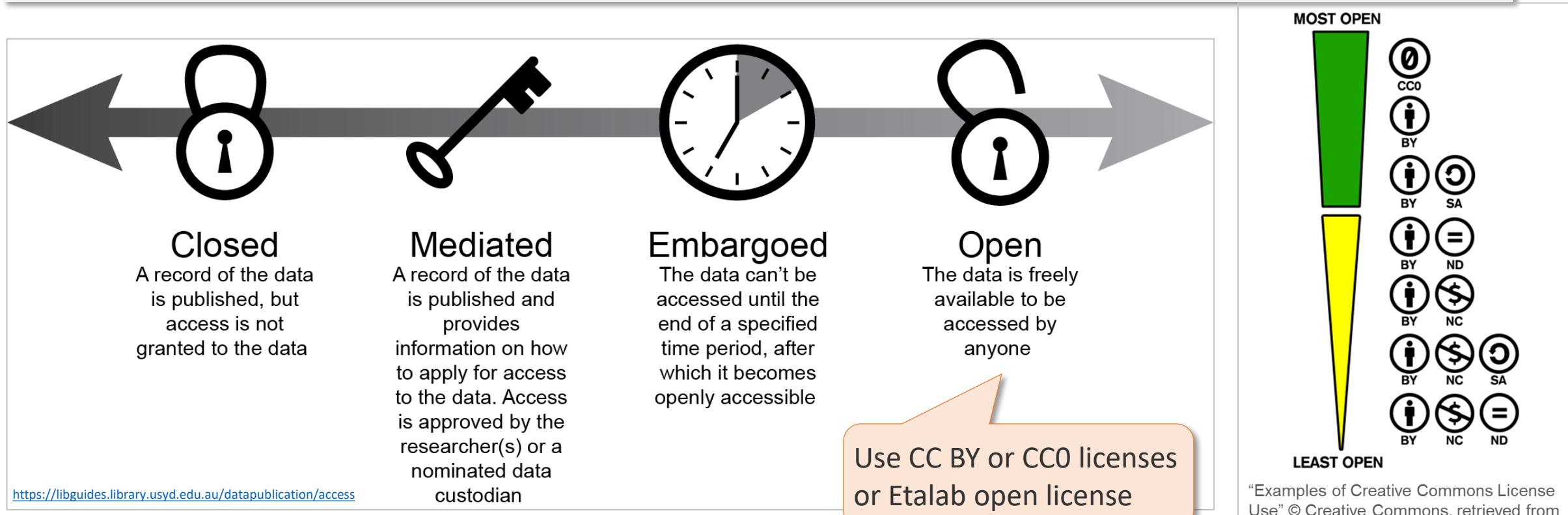
4. With whom can you share your OHM data?



"So-DRIIHM" Session

5. Data producers will always retain control over the dissemination of their data with different levels possible

Publishing your data doesn't necessarily mean that your data must be available to anyone and everyone
 → you will choose to apply specific restrictions on how people can gain access to your data

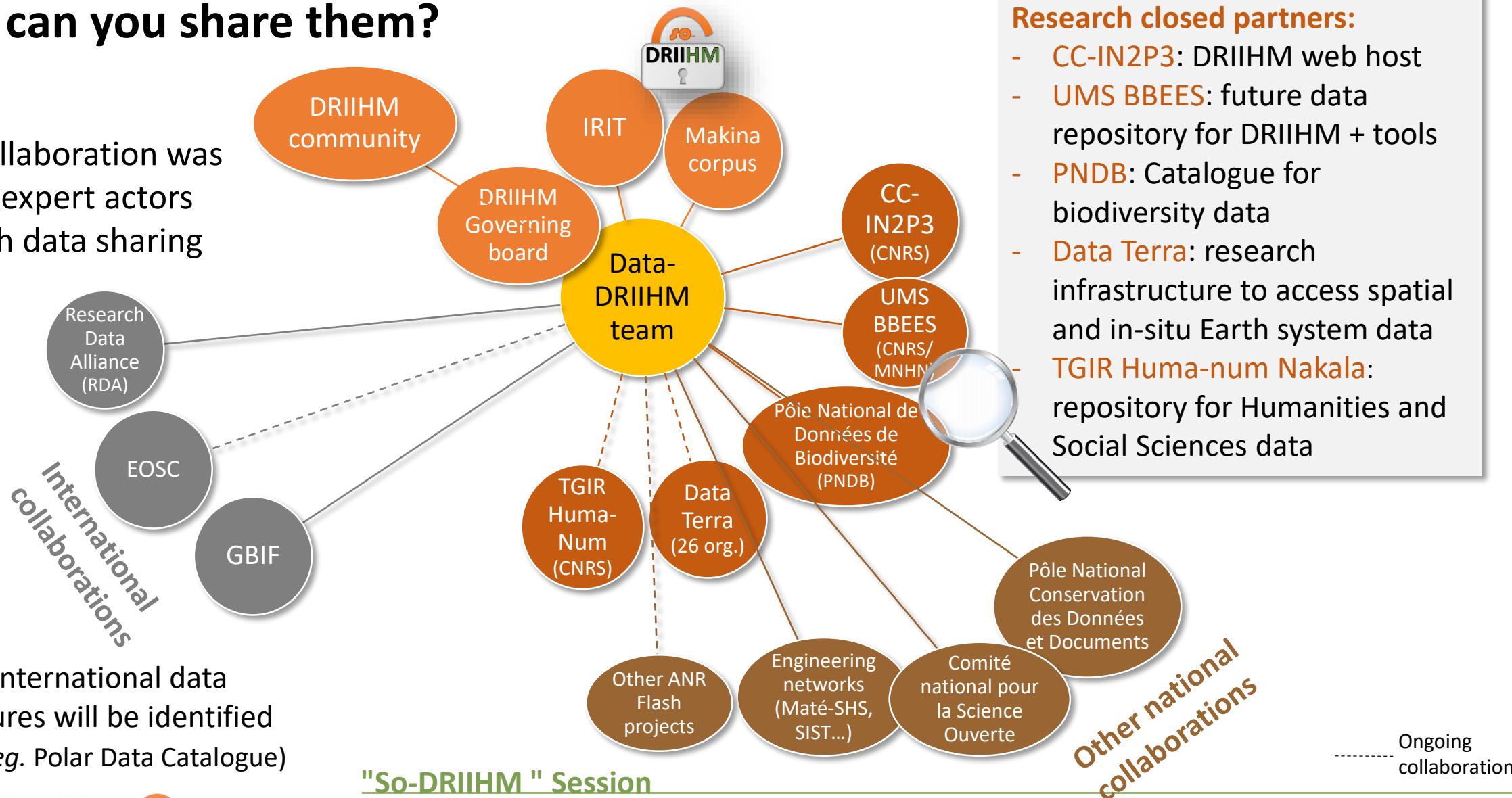


"Examples of Creative Commons License Use" © Creative Commons, retrieved from <https://creativecommons.org/examples/>, used under CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

6. How to save your OHM data?

Where can you share them?

A close collaboration was built with expert actors in research data sharing



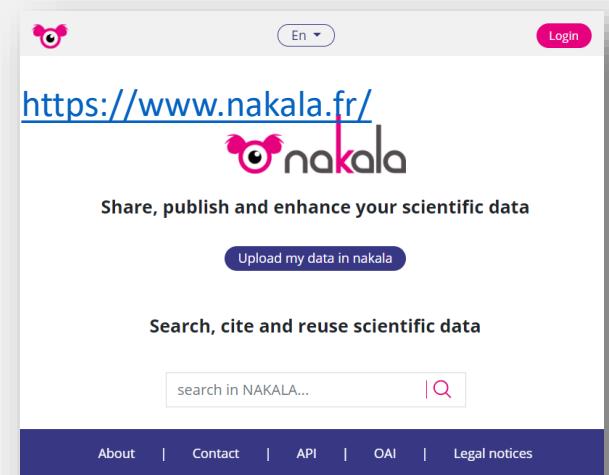
NB: Other international data infrastructures will be identified for OHMi (eg. Polar Data Catalogue)

7. The DRIIHM will not offer physical data storage but will direct to recommended data repositories

→ Several data repositories are currently under study:

- French thematic or institutional data repositories:
 - ✓ Data-InDoRES (UMS BBEES / CC-IN2P3 host) for CNRS-INEE (still under dev.)
 - ✓ PNDB for ecological data
 - ✓ Nakala (TGIR Human-Num) for Humanities and Social Sciences data
 - ✓ Data Terra for spatial and in-situ Earth system data
 - ✓ data.inrae.fr, dataverse.ird.fr, dataverse.cirad.fr, ...
- International thematic repositories:
 - ✓ PANGAEA for Earth & Environmental Science data
 - ✓ GBIF for biodiversity data
 - ✓ NCBI (Genbank...) for genetic/genomic data
- (Inter)national generalist repositories:
 - ✓ France: data.gouv.fr / geo.data.gouv.fr / geocatalogue.fr / recherche.data.gouv.fr
 - ✓ Europe: data.europa.eu and/or future EOSC
 - ✓ OpenAIRE and/or Zenodo

A complete list of recommended data repositories will be provided



Source: SO-DRIIHM 2021 survey + data-DRIIHM analyses

8. Among the data produced in the OHMs, some has been already shared (1/2)

- ✓ Data linked to published articles
- ✓ Data of common interest: data from weather/hydro. stations, land use data, biodiversity data, demographic data, ...
- ✓ Data considered shareable: cleaned raw data or processed/derived data
- ✓ Data you want to save/share/open!

PANGAEA.
Data Publisher for Earth & Environmental Science

Not logged in + ↗

SEARCH SUBMIT HELP ABOUT
CONTACT

[DOI:10.1594/PANGAEA.922437](#)

Citation: Arnaud, Fanny; Sehen Chanu, Lalandy; Grillot, Jules; Riquier, Jérémie; Piégay, Hervé; Roux-Michollet, Dad; Carrel, Georges; Olivier, Jean-Michel (2020): Cartographic and topo-bathymetric archive dataset on the French Rhône River (17th-20th centuries). PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.922437>

Rechercher dans NAKALA... | ↗
Fr Se connecter

Analyse du degré d'acceptabilité de la création d'une plateforme de dépôt et de partage des jeux de données des OHM (Observatoires Hommes-Milieux)
ID : 10.34847/nkl.3f777nrt Publié

Auteurs : Emilie Lerigoleur, Sarah Josse et Aurore Defays
[DOI:10.34847/nkl.3f777nrt](#)

Enquête par questionnaire en ligne réalisée en 2018 auprès de la communauté scientifique DRIIHM (Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieux) visant à évaluer le degré d'acceptabilité de la création d'une plateforme de dépôt et de partage des jeux de données des OHM (Observatoires Hommes-Milieux)
Le questionnaire est anonyme.
Le tableau de données correspond au...

Voir plus ↘

LabEx DRIIHM

Fichier	Visualisation
DRIIHM_enq...ults.csv	Afficher 250 entrées Chercher

Not logged in + ↗

SEARCH SUBMIT HELP ABOUT
CONTACT

[DOI:10.1594/PANGAEA.905838](#)

Citation: Rosset, Thomas (2019): Daily precipitation at Bernadouze. PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.905830>, In: Rosset, T (2019): High frequency biogeochemical and hydrological monitoring in two mountainous peatlands. PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.905838>

8. Among the data produced in the OHMs, some has been already shared (2/2)

Data linked to:

Hervé, V. and Lopez, P.J. (2020), Analysis of interdomain taxonomic patterns in urban street mats. Environ Microbiol, 22: 1280-1293. <https://doi.org/10.1111/1462-2920.14933>



Example

“The raw sequence data have been deposited in the NCBI Sequence Read Archive under the BioProjects PRJNA546091 (16S rRNA genes) and PRJNA316490 (18S rRNA genes)”

NCBI Resources How To

BioProject BioProject PRJNA546091 Search Help

Display Settings: Send to: Related information

biofilm metagenome Accession: PRJNA546091 ID: 546091 BioSample

Bacterial diversity in the street gutters of Paris, France SRA

Molecular inventory of the bacterial diversity colonizing the biofilms present in the street gutters. Taxonomy

Accession	PRJNA546091
Data Type	Targeted loci environmental
Scope	Environment
Organism	biofilm metagenome [Taxonomy ID: 718308] unclassified sequences; metagenomes; ecological metagenomes; biofilm metagenome
Submission	Registration date: 4-Jun-2019 Max Planck Institute for Terrestrial Microbiology
Related Resources	• MetaSUB & MetaSTREET
Relevance	Environmental

Project Data:

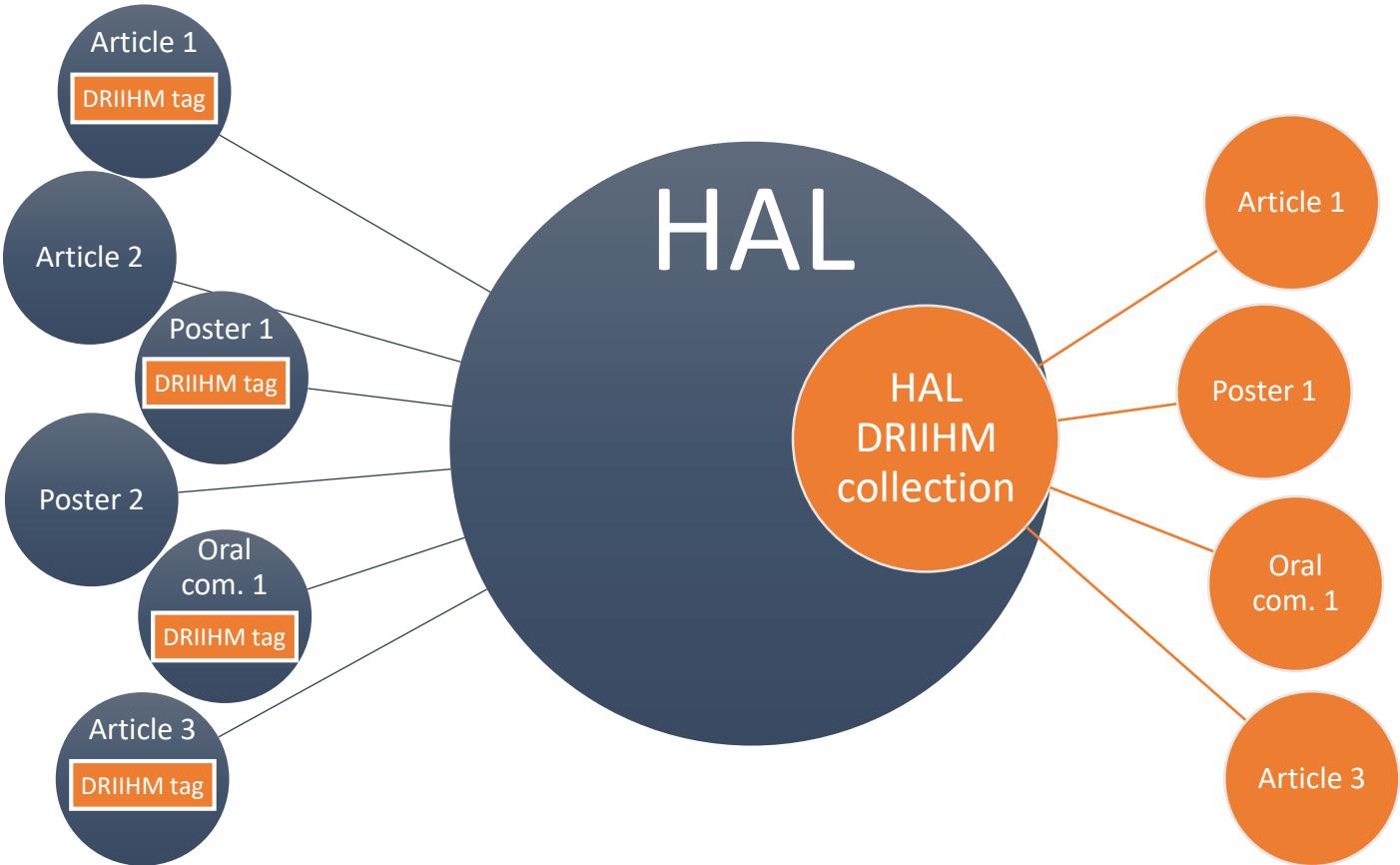
Resource Name	Number of Links
SEQUENCE DATA	60
SRA Experiments	

<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA546091>

Source: SO-DRIIHM 2021 survey

9. Towards non-dispersion of data and publications (1/3)

OHM
scientific
publications



HAL submission step

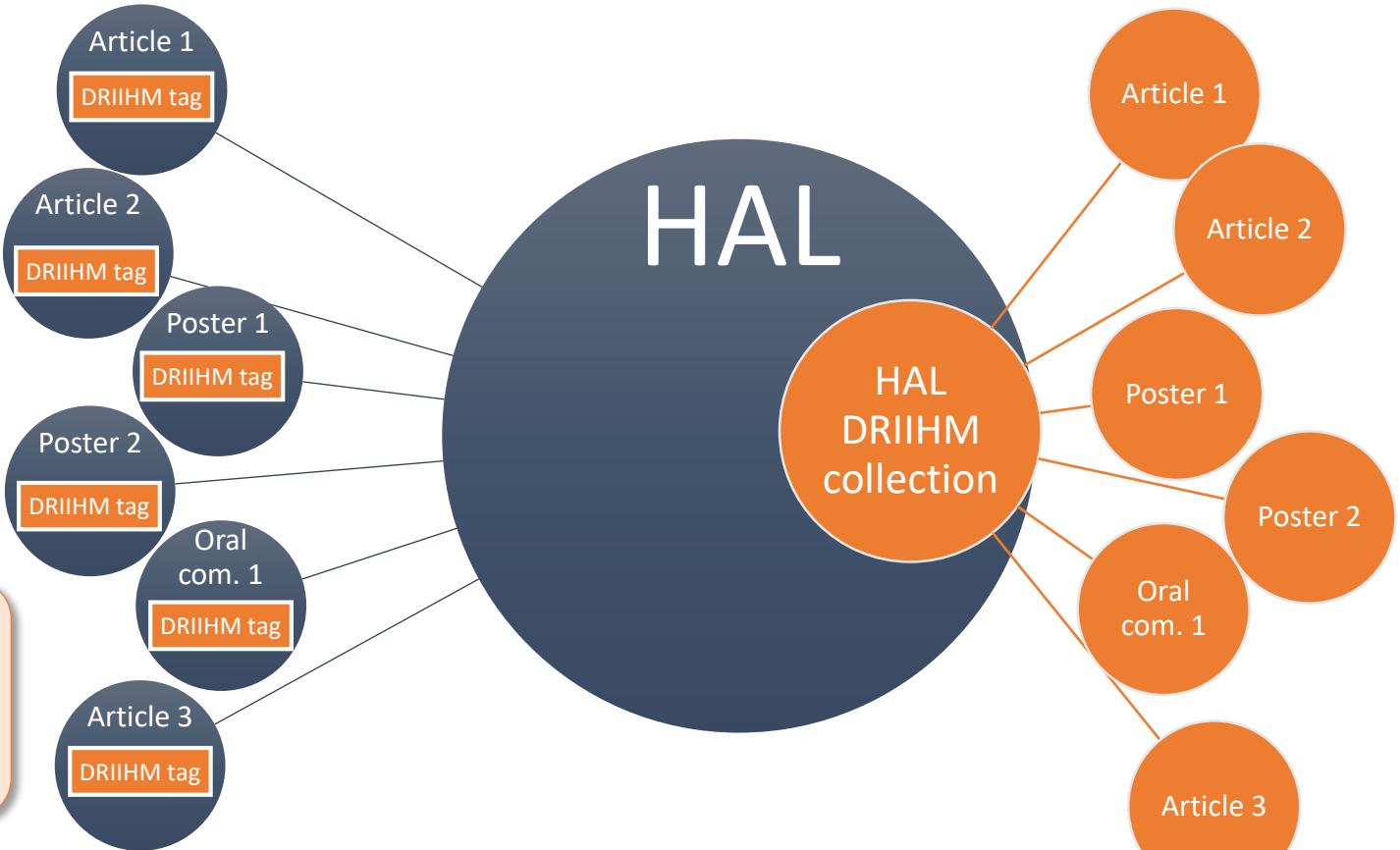
HAL searching for DRIIHM publications

"So-DRIIHM " Session

9. Towards non-dispersion of data and publications (2/3)

OHM
scientific
publications

use HAL and link each
publication to the
DRIIHM HAL collection

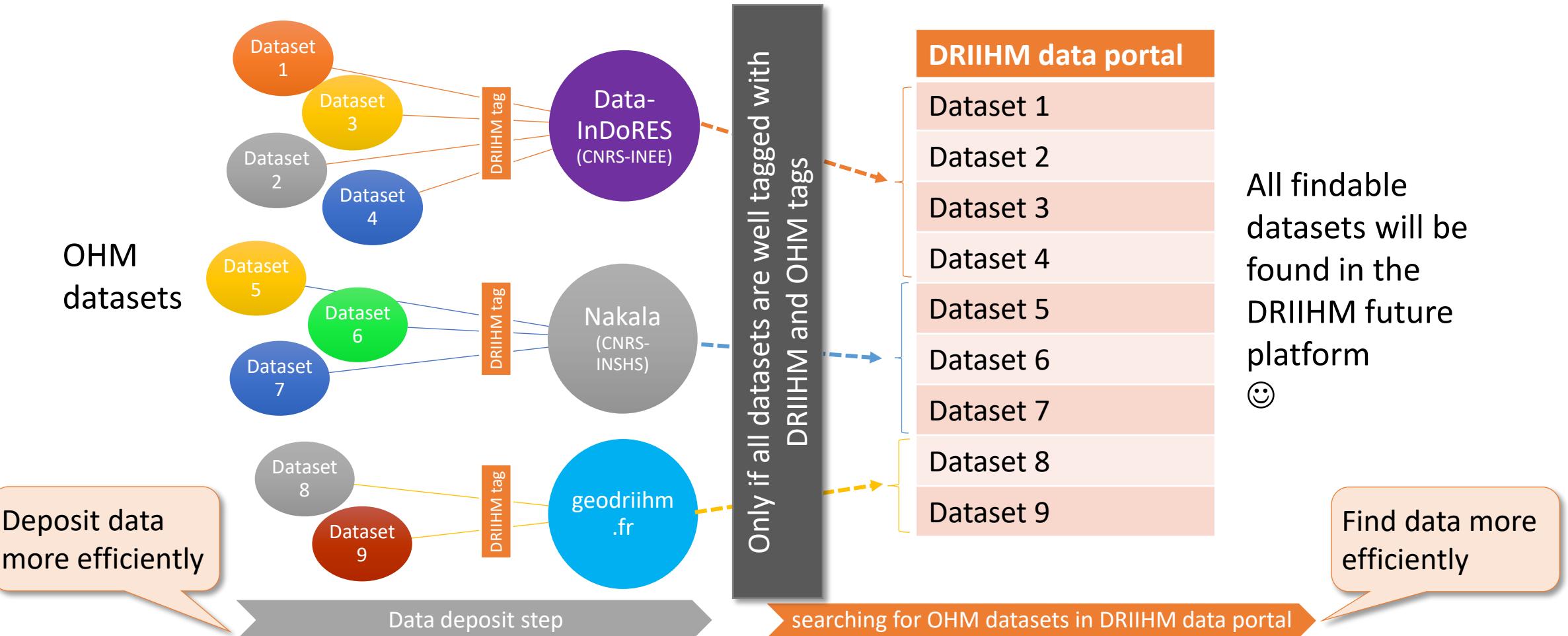


HAL submission step

HAL searching for DRIIHM publications

"So-DRIIHM " Session

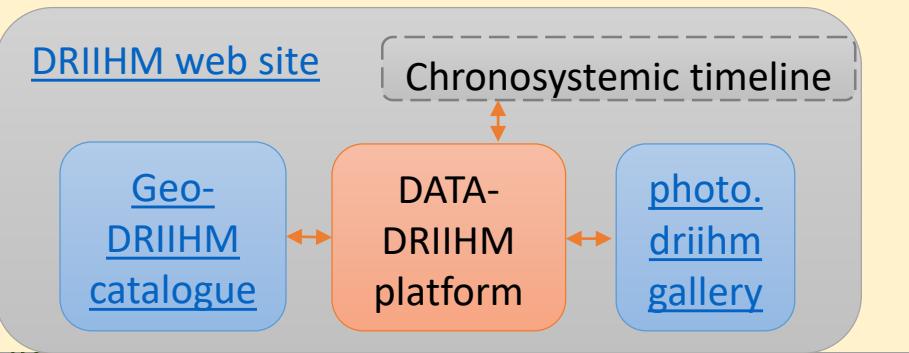
9. Towards non-dispersion of data and publications (3/3)



10. In SO-DRIIHM we imagine the future DATA-DRIIHM platform as an open science gateway

ID card

- **Name:** DATA-DRIIHM
- **Language:** english (first)
- **Web host:** [CC-IN2P3](#)
- **Technologies:** Django (python) / vue.js (javascript) / sparQL (semantic web) based on DCAT (= an RDF vocabulary designed to facilitate interoperability between Web data catalogues)
- **2 levels of web access:**
 - Open access (public site)
 - Restricted access for OHM participants
- Integrated in the DRIIHM web site



An **open science gateway** to guide you to:

- **Search for OHM data and other linked data** (land use, meteo. data, demographic data...)
- **Search for (OHM) publications**
- **Deposit data in a recommended data repository:**
 - choose the data repository
 - choose the access & use license
 - tag correctly data and publications
- **Learn how to make more FAIR your data**
- **Write a Data Management Plan**
- **Draft a data paper**
- **Cross-analyze selected datasets and visualize them**

PRIORITY

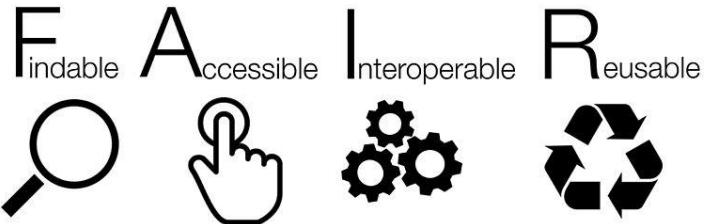
11. SO-DRIIHM vision... in summary

1. A step towards sharing and opening up OHM data, but need for support and consideration of potential associated risks
2. The ANR SO-DRIIHM project encourages the sharing and opening of data within the OHMs, between them and beyond
3. The Data-DRIIHM team with IRIT and Makina corpus at the helm
4. You can share your data within the OHMs, between them and beyond
5. Data producers will always retain control over the dissemination of their data with different levels possible
6. A close collaboration was built with expert actors in research data sharing
7. The DRIIHM will not offer physical data storage but will direct to recommended data repositories
8. Among the data produced in the OHMs, some has been already shared
9. Towards non-dispersion of data and publications
10. In SO-DRIIHM we imagine the future DATA-DRIIHM platform as an open science gateway

"So-DRIIHM " Session

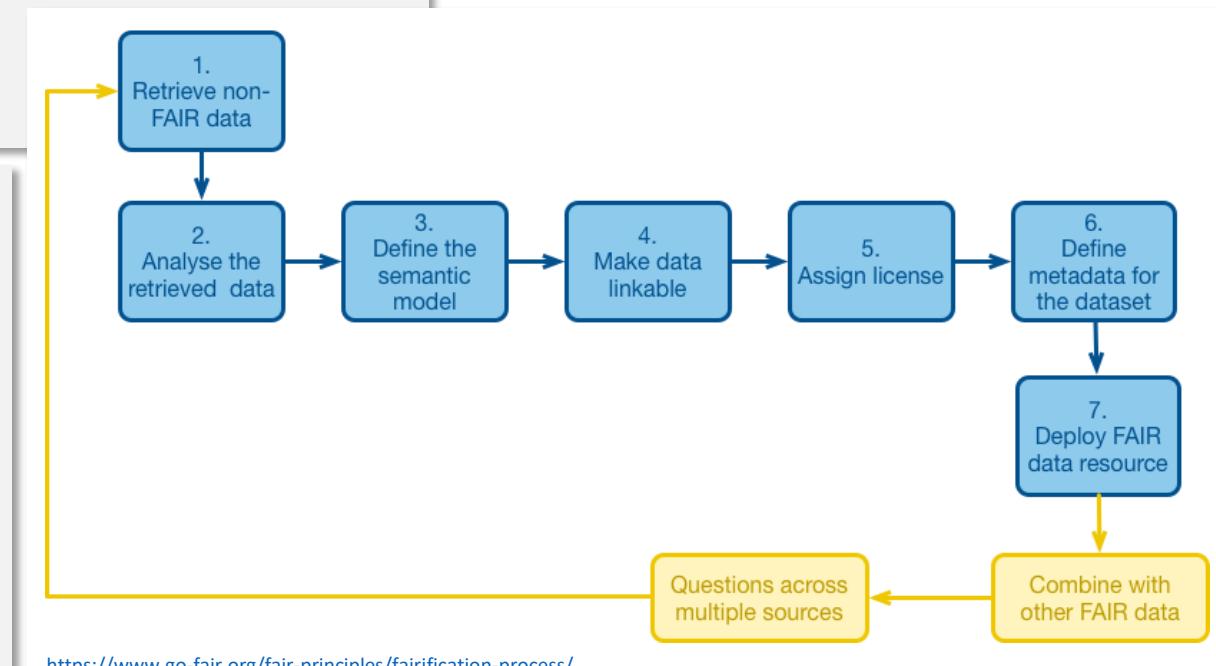
1. In SO-DRIIHM we support the open science approach through OHM pilot projects (1/2)

- **Trafic** (Fondation de France - OHM Littoral Caraïbe)
- **PastServ** (APR 2020 - OHM Haut Vicdessos)
- **Evol. sédim. Delta du Rhône** (APR 2020 - OHM Vallée du Rhône)
- **Base de données RhônEco** (OHM Vallée du Rhône)
- **GéOyapock** (APR 2020 - OHM Oyapock)
- **Post-doc** (APR 2020 - OHMi Téssékéré)



→ The SO-DRIIHM team helps researchers to carry out the “FAIRification” process:

- ✓ Diagnose data management and/or write a Data Management Plan (DMP)
- ✓ Identify which data to preserve/share/publish
- ✓ Check for (non-)compliance with the GDPR/legal constraints (only if personal data)
- ✓ Structure and document data using standards
- ✓ Choose the data repository and deposit data
- ✓ Write a data paper (if desired)



1. In SO-DRIIHM we support the open science approach through OHM pilot projects (2/2)

- **Trafic** (Fondation de France - OHM Littoral Caraïbe)
- **PastServ** (APR 2020 - OHM Haut Vicdessos)
 - ✓ A specific dataset has been identified to be preserved and published
 - ✓ We will make it as FAIR as possible
- **Base de données RhôneEco** (OHM Vallée du Rhône)
 - ✓ Diagnose the database and data management
 - ✓ Initial recommendations and support:
 - write a DMP using [DMP opidori tool](#)
 - identify which part of the database will be shared/open
 - create a data dictionary to improve reuse
- **GéOyapock** (APR 2020 - OHM Oyapock)
 - ✓ Standardize as possible an online survey based on past surveys
 - ✓ Start the workflow design of a future open source web application
- **Post-doc** (APR 2020 - OHMi Téssékéré)
 - ✓ Check sensitive data management and GDPR legislative constraints
 - ✓ Identify datasets to make as FAIR as possible

Focus on TRAFIC project results
→ see Jung *et al.* next presentation!

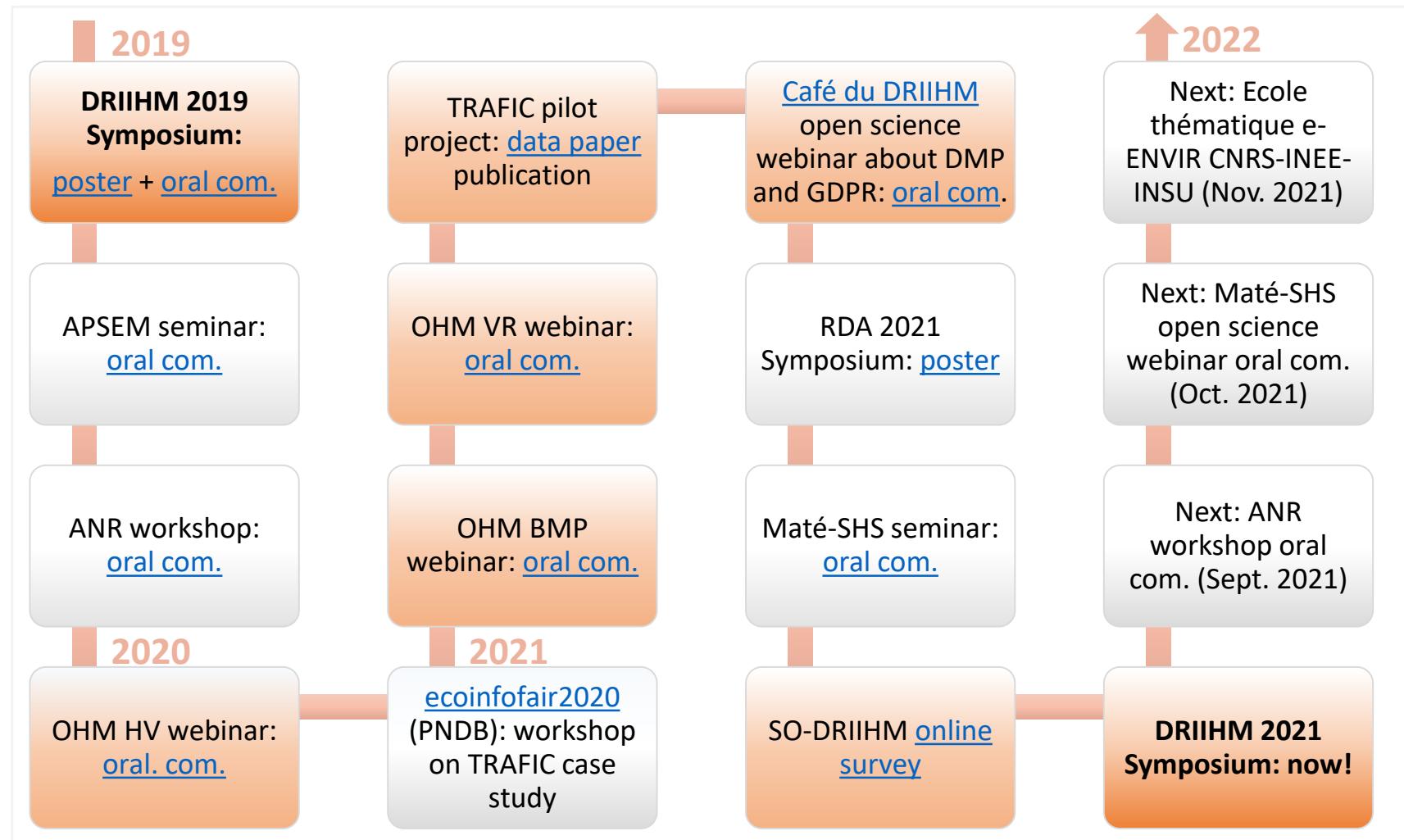
2. In SO-DRIIHM we communicate, we inform, we train

✓ Several interventions in DRIIHM internal and external seminars

- ✓ Topics
 - ✓ Open Science context & best practices
 - ✓ SO-DRIIHM project
 - ✓ Results of surveys

+ Several participations to pilot project meetings

All publications are referenced in HAL
 → [SO-DRIIHM HAL collection](#) ☺



"So-DRIIHM" Session

3. Survey first results validate the need for methodological and technical support



Load unfin

A summary will be provided later

Share your experience of data sharing!

DRIIHM - Interdisciplinary Research Device on Human-Environment Interactions

We are asking you because you are involved in at least one of the Human-Environment Observatories (OHM) of the LabEx DRIIHM.

Data sharing is a major challenge for the scientific community today. Sharing is both making available the data we produce, but also accessing data useful to our research.

We are currently designing a web platform at the DRIIHM level in order to facilitate the sharing of research data and to accompany you in this process. This project is led by the engineers/researchers of the Data-DRIIHM group within the framework of the DRIIHM ANR Open Science project (SO-DRIIHM).

It's obviously essential that this platform adapts to your practices and meets your needs. To this end, we invite you to share your experience of data sharing by answering the questionnaire via the following link:
<https://enquetes.univ-tlse2.fr/index.php/968188?lang=en>

Your participation requires about 15 minutes. It is essential to the success of the project!

Deadline to answer: July 9th!

We warmly thank you for your answers and for your involvement in the service of the collective.

Sincerely
The DATADRIIHM group

There are 34 questions in this survey.

To continue, please first accept our questionnaire p
[Show policy](#)

You can still participate!

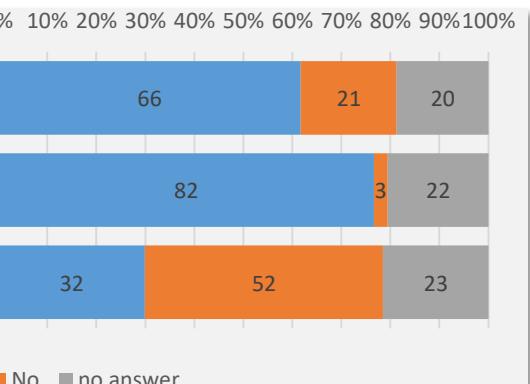
English version: <https://enquetes.univ-tlse2.fr/index.php/968188?lang=en>

French version: <https://enquetes.univ-tlse2.fr/index.php/968188?lang=fr>

Do you use external data?

Do you share* your data?

Do you deposit your data?



* Share data mainly with OHM project partners, lab colleagues and OHM collaborators

4. First IT technical tasks concerned the design of the DATA-DRIIHM search module

✓ The web platform has started its conception & development (mainly back-end, then front-end)

✓ IT working group

- ✓ SO-DRIIHM subgroup (GEODE/IRIT/ECCOREV)
+ 6-month intern (March-Sept. 2021)
- ✓ Closed collaboration with the [datanoos](#) network coordinated by IRIT (co-developments by interns)

✓ First IT technical tasks performed

- ✓ Creation of a DCAT web semantic database containing first the géoDRIIHM catalogue metadata
- ✓ Design mock-ups of the future web platform (Search and Deposit IT modules)
- ✓ First web developments → a first proof of concept (POC) designed (not accessible)

We will discover & analyze these mock-ups during our **SO-DRIIHM workshop**
→ just after this presentation!

5. SO-DRIIHM next steps



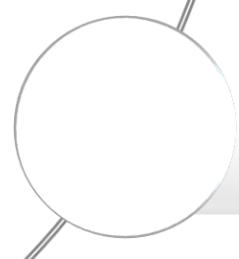
Continue to support OHM pilot projects
Support DRIIHM transversal projects



Continue to communicate about open science and best practices & tools
ANR workshop oral com. (Sept. 2021) + Maté-SHS open science webinar oral com. (Oct. 2021)
+ Ecole thématique e-ENVIR CNRS-INEE-INSU (Nov. 2021) + ...



Continue to build and reinforce collaboration with (inter)national data research infrastructures
Towards a DRIIHM Open Science policy...



Continue to co-design and develop the interoperable DATA-DRIIHM platform
We will need beta testers!

"So-DRIIHM " Session

Thank you!

CONTACT US

Data-DRIIHM team:

Fanny Arnaud, Armelle Decaulne, Priscilla Duboz, Samuel Dunesme, Arnaud Jean-Charles, Iwan Le Berre, Emilie Lerigoleur, Mathieu Massaviol, Kristell Michel, Bertrand Morandi, Corinne Pardo, Jean-Claude Raynal, Grzegorz Skupinski, Marie-Laure Trémélo

Contact:

emilie.lerigoleur@univ-tlse2.fr

ORCID: [0000-0002-0864-659X](https://orcid.org/0000-0002-0864-659X)

IRIT:

Pascal Dayre

COMPLEMENTARY RESOURCES

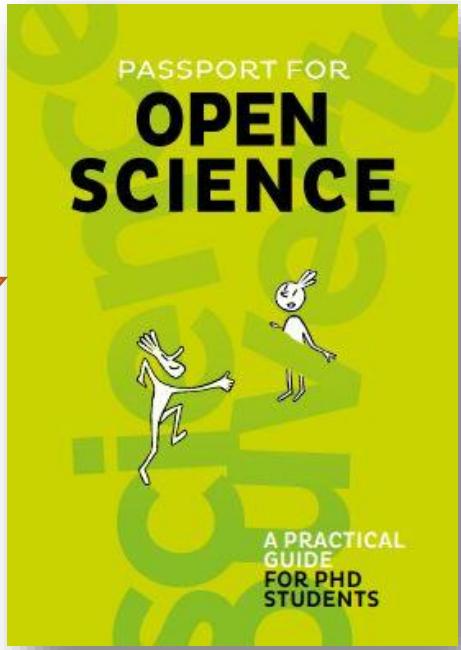
Some best practice guides and tutorials

<https://www.ouvrirlascience.fr>

<https://doranum.fr>

[Passport for open science: a practical guide for PhD students](#)

[Guide de bonnes pratiques sur la gestion des données de la recherche \(CNRS-MITI\)](#)



[English](#) and [French](#) versions

"So-DRIIHM " Session

PNSO2 encourages the journals' editorial committees and scientific editors to request that the data and code associated with the texts submitted to be provided

Example: **NATURE - Scientific Data** journal asks authors to submit datasets to an appropriate [public data repository](#)

→ cf. [Data Repository Guidance](#)



Social sciences →

Archaeology Data Service	view re3data entry
Harvard Dataverse	view re3data entry
ICPSR	view re3data entry
Open Science Framework	view FAIRsharing entry
Qualitative Data Repository	view FAIRsharing entry
UK Data Service	view re3data entry

"So-DRIIHM" Session

Earth, Environmental and Space sciences →

Broad scope Earth & environmental sciences →

NASA Goddard Earth Sciences Data and Information Services Center	view re3data entry
NERC Data Centres	view re3data entry
PANGAEA	view re3data entry
National Tibetan Plateau/Third Pole Environment Data Center	view FAIRsharing entry
NOAA National Centers for Environmental Information (DOIs only assigned to deposited data on request)	view re3data entry
HydroShare (CUAHSI)	view FAIRsharing entry

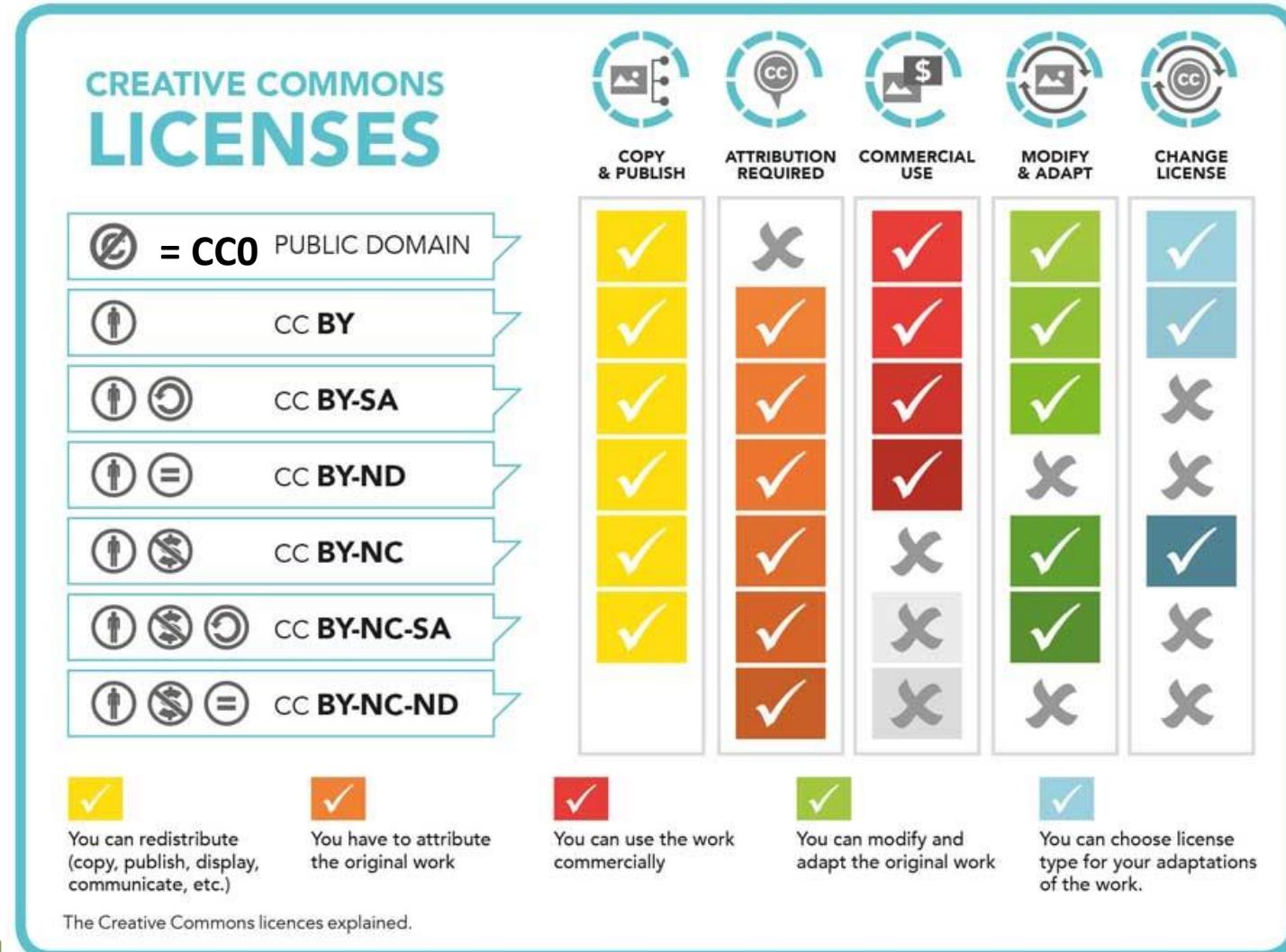
Ecology →

TERN Data Discovery Portal	view FAIRsharing entry
Environmental Data Initiative (formerly LTER Network Information System Data Portal)	view re3data entry
Global Biodiversity Information Facility (GBIF)	view FAIRsharing entry
KNB: The Knowledge Network for Biocomplexity	view FAIRsharing entry

License to share: How the Creative Commons licensing system encourages the remixing and reuse of published materials

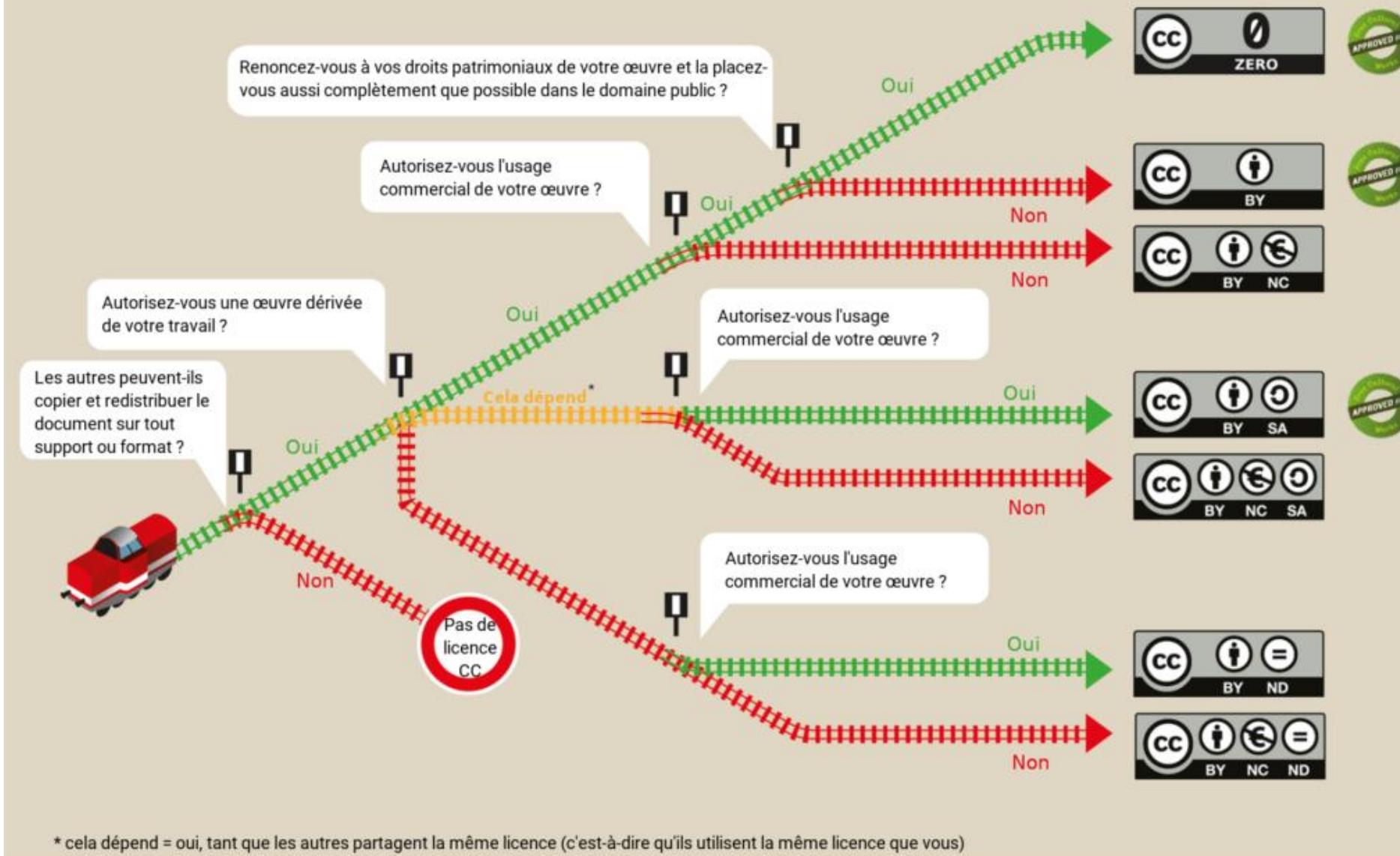


Source: <https://researchoutreach.org/>



"CHOO-CHOO-CHOOSE YOUR LICENSE!"

ANNEX



* cela dépend = oui, tant que les autres partagent la même licence (c'est-à-dire qu'ils utilisent la même licence que vous)



Ce graphique « Choo-Choo-Choose your license » est basé sur le travail « Welche CC-Lizenz ist die richtige für mich ? » réalisé par Barbara Klute et Jörn Muuß-Merholz pour <http://www.wb-web.de> sous la licence CC BY-SA 3.0. La version française est une traduction de Cédric Frayssinet sous la même licence.