



Science ouverte et OHM :

Le projet SO-DRIIHM pour faciliter le partage des données des OHM



Emilie Lerigoleur (GEODE - OHM Pyrénées Haut Vicdessos)
pour le collectif SO-DRIIHM





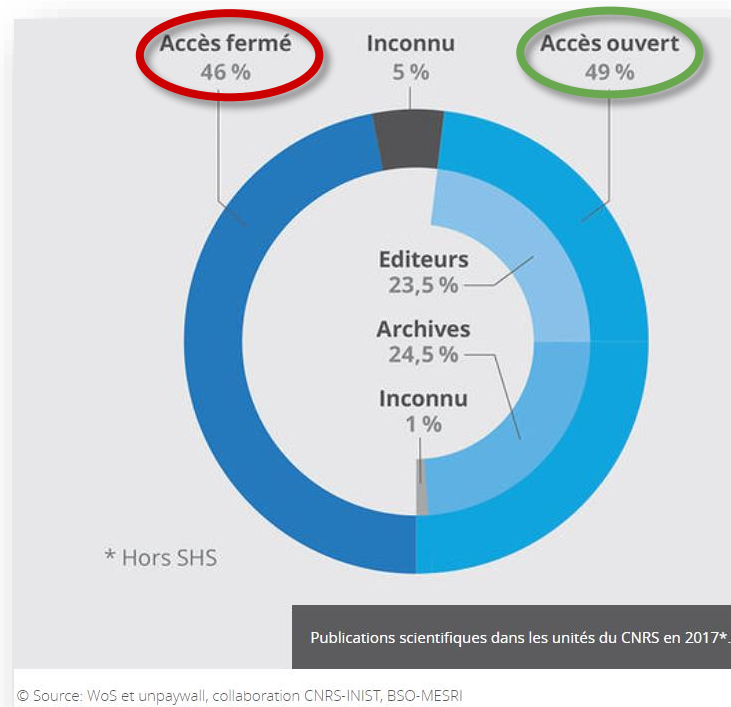
**1. Etat des lieux sur le
partage des données au sein
des OHM**

**2. Contexte Science Ouverte
en pleine r()évolution**

**3. Projet ANR SO-DRIIHM
2020-2023**

1. Etat des lieux sur le partage des données au sein des OHM

- Constat 1: dispersion et faible accessibilité des publications scientifiques
 - Référencement dans archive ouverte HAL devenu obligatoire va améliorer la situation
 - Pas toujours de citation du LabEx DRIIHM (cf. charte OHM)
 - Rappel [consignes dépôt HAL](#)



CCSD HAL Episciences.org Sciencesconf.org Support Connexion

OHM Vallée du Rhône

Page d'accueil Consultation des dépôts Recherche

Recherche par documents

Filtres: Tous x

187 résultats

enregistrer la recherche

TYPE DE DOCUMENT

- Communication dans un congrès (98)
- Article dans une revue (56)
- Poster (12)
- Chapitre d'ouvrage (10)
- Rapport (5)
- Direction d'ouvrage, Proceedings, Dossier (3)
- Thèse (2)
- Vidéo (1)

AUTEUR

Filtrer

- P. Allard (5)

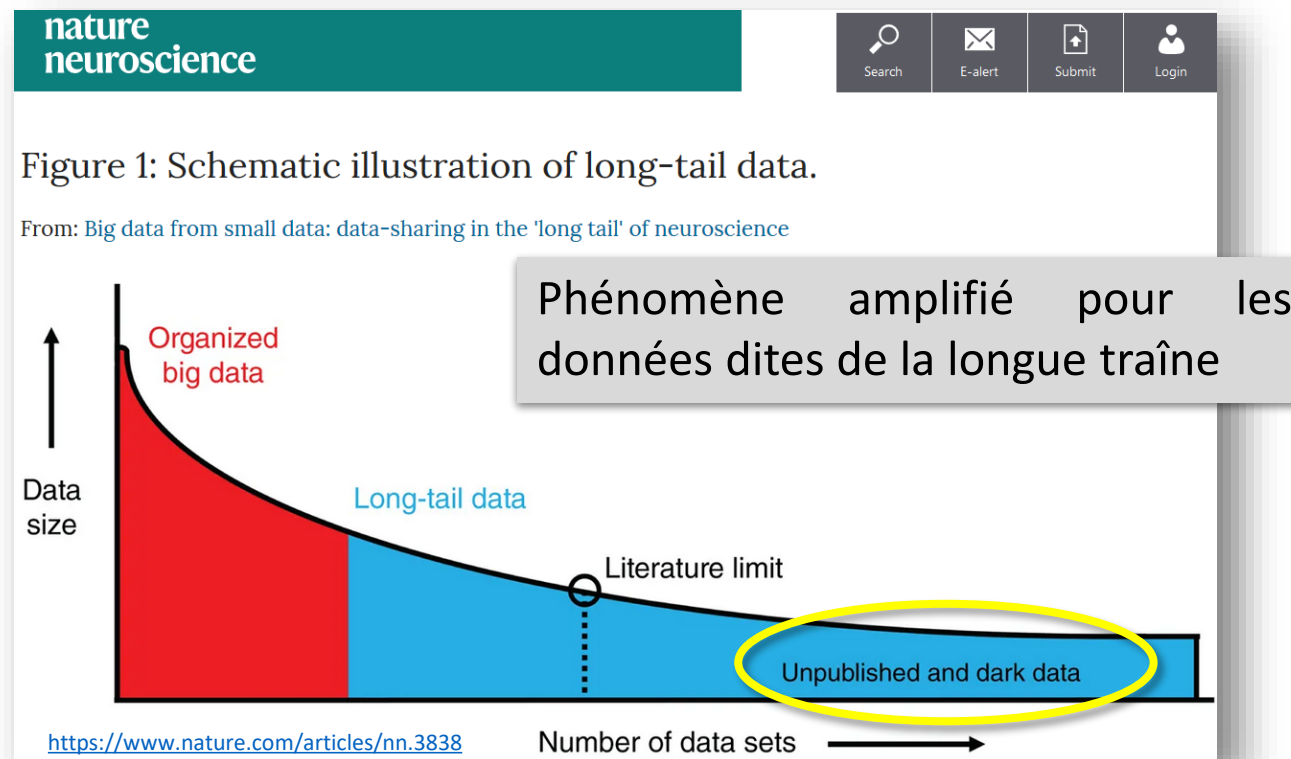
hal-02966972v1 **Communication dans un congrès**
Fanny Arnaud, Mathieu Cassel, Guillaume Fantino, Daniel Vazquez-Tarrio, Anne Clemens *et al.*
Atelier "Les traceurs sédimentaires pour comprendre la trajectoire morphologique d'un cours d'eau"
Pêche aux outils scientifiques, GRAIE-ZABR, Oct 2020, Miribel, France

hal-02954094v1 **Poster**
Silvia Flaminio. **Loyettes, an Unbuilt Dam: Breaking up the Rhône River's Geohistory**
International Symposium of LabExDRIIHM, Oct 2019, Lyon, France. 2019

hal-02954093v1 **Poster**
A. Morereau, C. Begorre, H. Lepage, A. Dabrin, F. Eyrolle *et al.* **Anthropisation of the Rhône River during the last century: case of main pollutants measured in sediment cores**
International Symposium of LabExDRIIHM, Oct 2019, Lyon, France. 2019

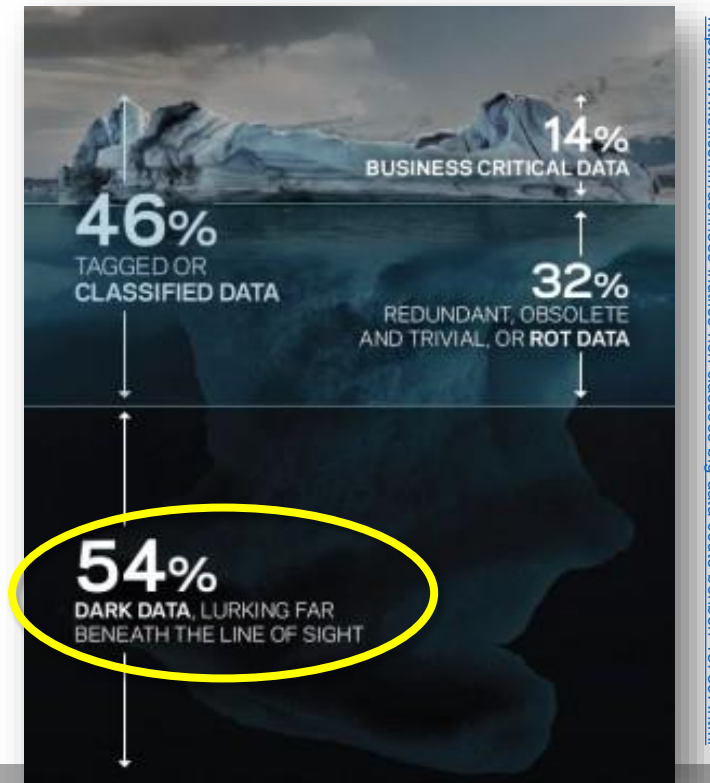
1. Etat des lieux sur le partage des données au sein des OHM

- Constat 2: dispersion, méconnaissance voire perte des données des OHM
 - données publiées “encapsulées” dans les PDF ou les “supplementary data” ou entrepôts inconnus



1. Etat des lieux sur le partage des données au sein des OHM

- Constat 2: dispersion, méconnaissance voire perte des données des OHM (suite)
 - données non publiées “fossilisées” = **dark data** : 10 % des données dans publications, “le reste demeurant disponible mais non utilisé sur les disques durs d’ordinateurs” (source : Durand-Barthez 2018)



1. Etat des lieux sur le partage des données au sein des OHM

- Constat 2: dispersion, méconnaissance voire perte des données des OHM (fin)
 - données **peu accessibles via les géocatalogues** des OHM (dont [ELVIS](#)) et du [géo-DRIIHM](#) central

The screenshot displays the ELVIS web interface. At the top, there is a search bar with the text "Rechercher ..." and a magnifying glass icon. Below the search bar, it indicates "Rechercher parmi 305 jeux de données, services et cartes, ...". The main content area is divided into two sections: "Parcourir par Thèmes" and "Types de ressource".

Parcourir par Thèmes

- Eaux intérieures (198)
- Société (55)
- Planification/Cadastré (3)
- Biote (2)
- Climatologie/Météorologie/Atm... (1)
- Environnement (112)
- Altitude (32)
- Structure (3)
- imageryBaseMapEarthCover (2)
- Agriculture (1)
- Imagerie/Cartes de base/Occu... (83)
- Informations géoscientifiques (29)
- Santé (3)
- Limites (1)

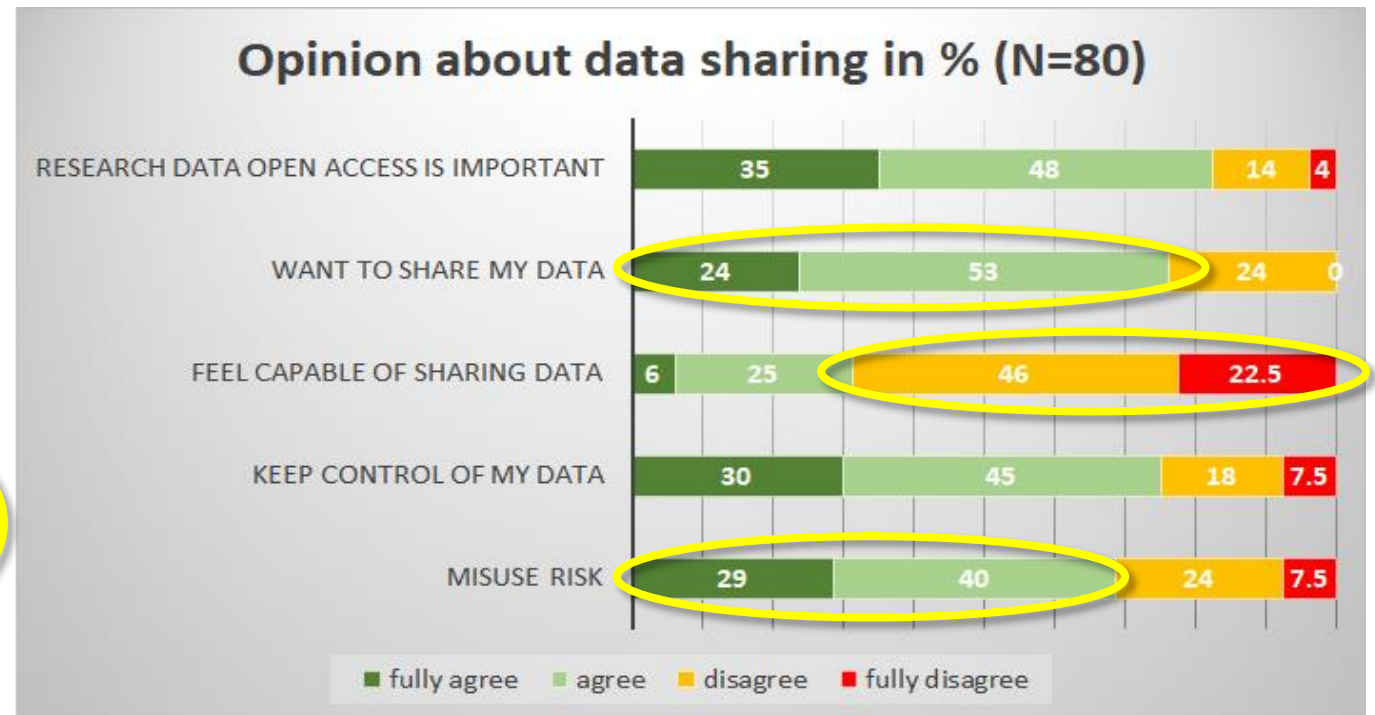
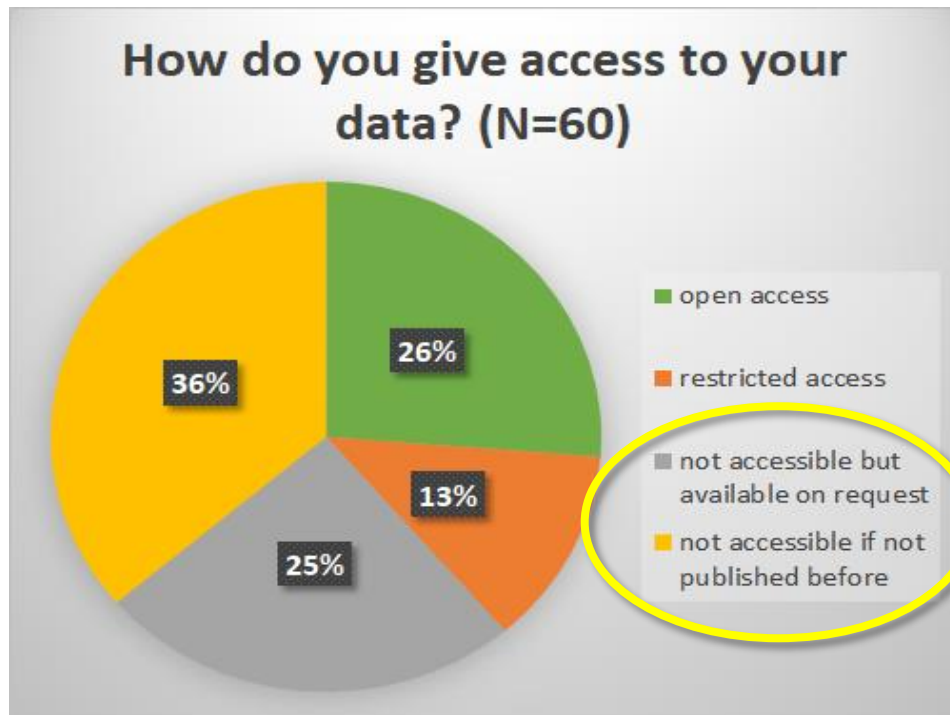
Types de ressource

- Jeu de données (291)
- Carte statique (11)
- Carte (9)
- Jeux de données non géograp... (2)

At the bottom of the interface, there are buttons for "Nouveautés" and "Les plus vues". Below these buttons, there are four preview cards for various data sets: "Cartes bathymétriques du Rhône de Lyon à la ...", "Photothèque de l'OHM VR (2018)", "Levé LiDAR sur le Vieux Rhône de Chautagne (...)", and "Emprise actuelle des casiers Girardon des Vieu...". The logo of the University of Lyon is visible in the bottom right corner of the interface.

1. Etat des lieux sur le partage des données au sein des OHM

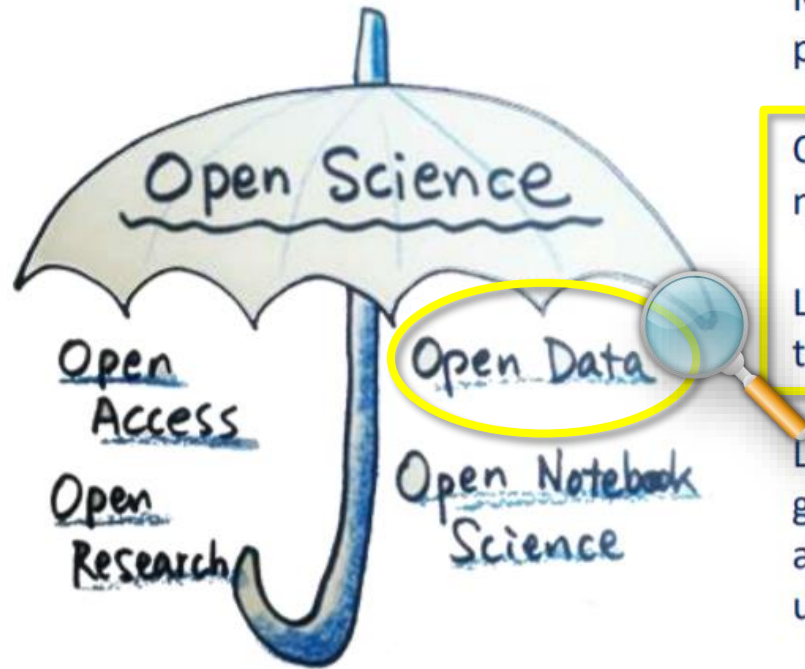
- Enquête 2018: analyse du degré d'acceptabilité de la création d'une plateforme de partage des jeux de données pour les OHM



→ Un pas vers le partage et l'ouverture des données des OHM, mais nécessité d'accompagnement (bonnes pratiques, outils facilitateurs) et prise en compte des potentiels risques associés

2. Contexte Science Ouverte en pleine (r)évolution

La science ouverte



Mouvement de diffusion gratuite de la production scientifique.

Ouvert autant que possible, protégé autant que nécessaire.

Lever les barrières de l'accès en maintenant toutes les protections du droit d'auteur.

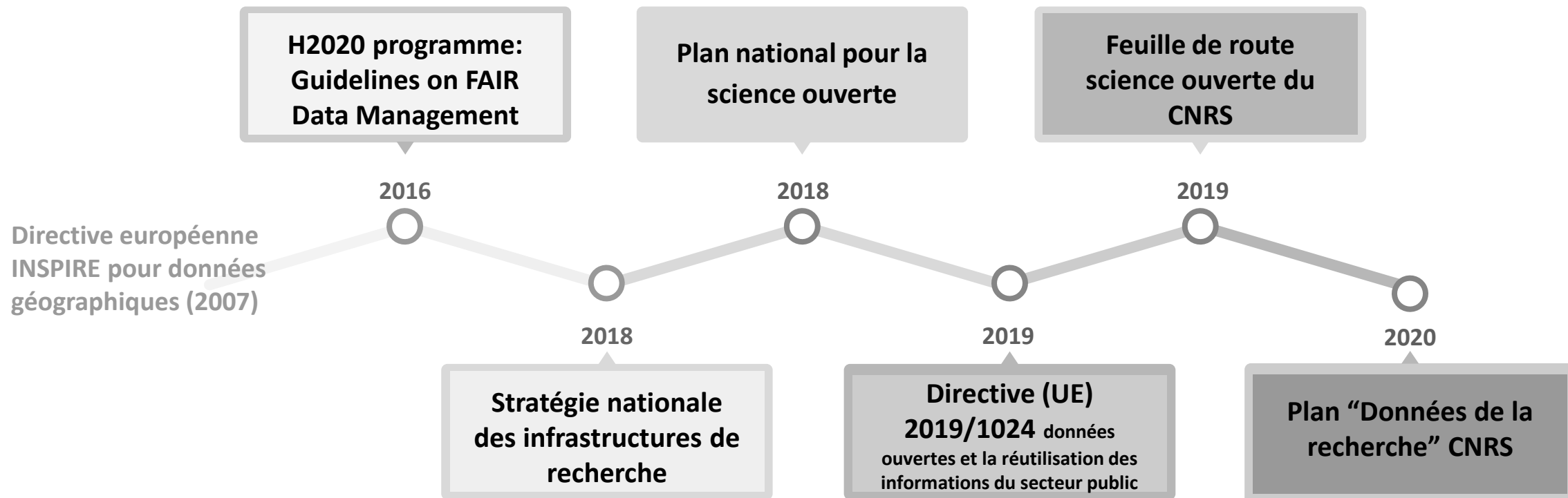
Le développement de la science ouverte doit être global. En France il ne peut se faire qu'en synergie avec l'ensemble des acteurs de la recherche : universités, organismes, écoles...

Intervention d'Antoine Petit, PDG du CNRS, aux Journées nationales de la Science Ouverte 2019

“La science ouverte, une révolution nécessaire”

2. Contexte Science Ouverte en pleine (r)évolution

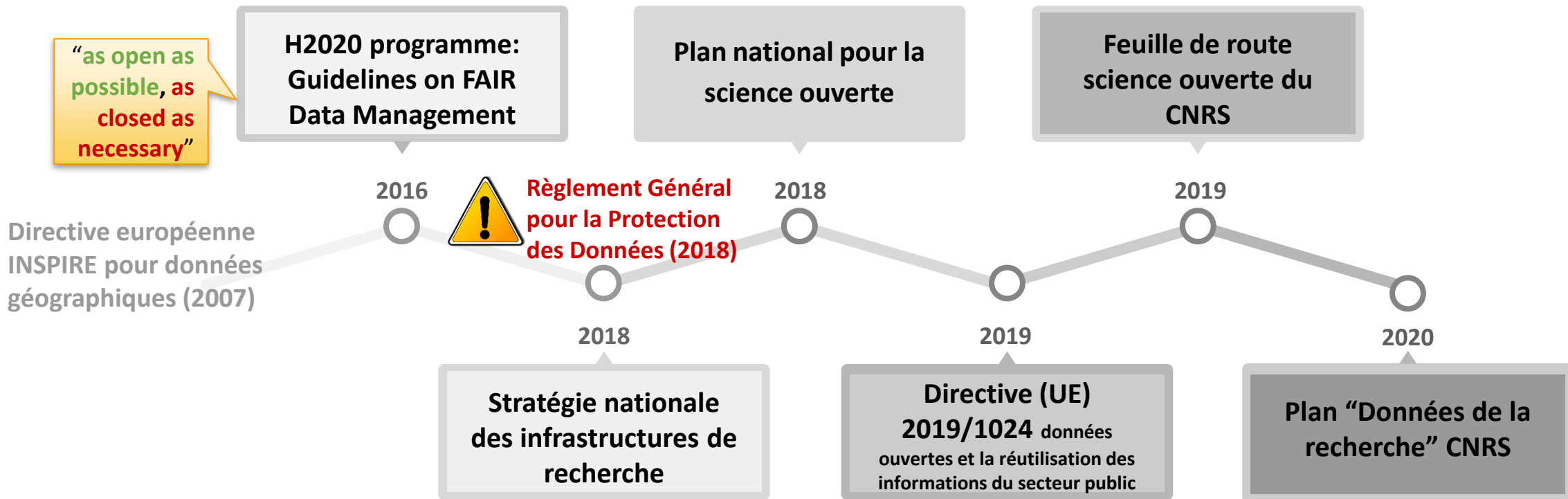
- Principaux textes de référence en lien avec l'ouverture des données de la recherche



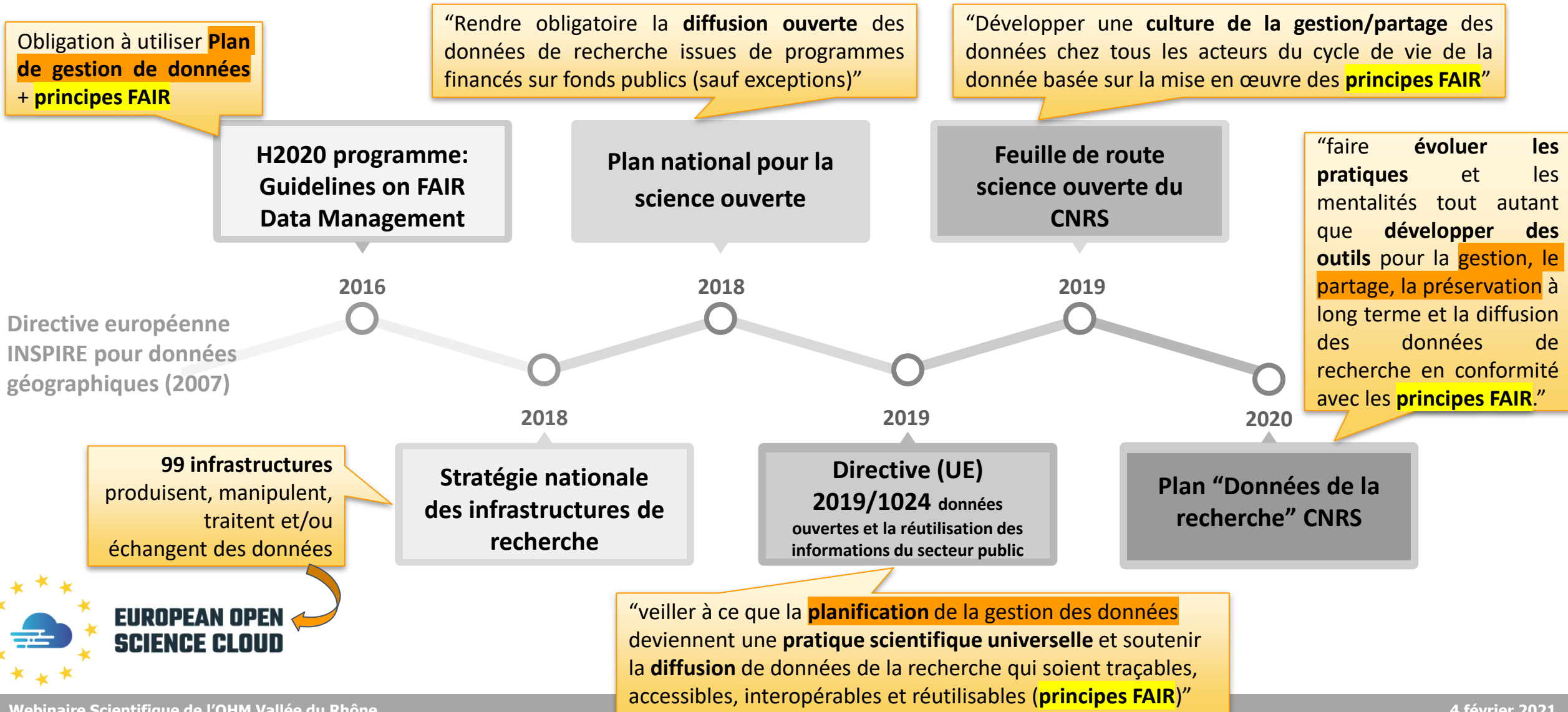
Rem : s'applique aux données **sur fonds publics > 50%** sauf **exceptions** légitimes encadrées par la loi (secret professionnel, secrets industriels et commerciaux, données personnelles ou contenus protégés par le droit d'auteur)

2. Contexte Science Ouverte en pleine (r)évolution

- Principe “**ouvrir les données autant que possible, les fermer autant que nécessaire**”



● 2 mots-clés : principes FAIR et Plan de Gestion de Données



2. Contexte Science Ouverte en pleine (r)évolution

● Principes FAIR : augmenter le potentiel des données

Exemple fictif

Tableau de résultats d'analyse de polluants

- Pourquoi ? **étude contamination**
- Quoi ? **liste polluants**
- Où ? **stations sur cours d'eau X**
- Quand ? **2017-2019**
- Comment ? **prélèvements d'eau selon fréquence 1 fois/sem, instruments de mesure, variables...**
- Par qui ? **Mr XY et Mme XX**
- Projet(s) ? Quel(s) OHM ? **OHM VR**
- Financement ? **DRIIHM**
- Format fichier ? **Excel**

(métadonnées existantes mais non formalisées, ie. dans la tête du chercheur !)



Rem: une donnée peut être FAIR sans être ouverte !

ID : DOI 10.5281/zenodo214689987
+ Métadonnées standardisées en ligne :
Tableau de résultats d'analyse de polluants

- Pourquoi ? **étude contamination**
- Quoi ? **liste polluants**
- Où ? **stations sur cours d'eau X**
- Quand ? **2017-2019**
- Comment ? **prélèvements d'eau selon fréquence 1 fois/sem, instruments de mesure, variables standardisées**
- Par qui ? **Mr XY et Mme XX (contacts)**
- Projet(s) ? Quel(s) OHM ? ...
- Financement ? **DRIIHM**
- Format fichier ? **Excel CSV ou ODS**
- **Accès restreint aux données, sur demande (cf. contacts) mais accès ouvert aux métadonnées**
- **Licence CC BY-NC-ND 2.0**

2. Contexte Science Ouverte en pleine (r)évolution

- **Plan de Gestion de Données - PGD** (*Data Management Plan - DMP*)

- Recommandé et désormais **exigé** par de nombreux **financeurs** dont l'Agence Nationale pour la Recherche
- A l'échelle d'un projet, système d'information, structure, thèse...
- Document **évolutif**, éditable collaborativement à l'aide par exemple de l'outil CNRS-INIST
- Offre une **transparence**, permet de se poser les bonnes questions et d'**anticiper** !



1. Data Collection and Documentation



- What kind of data are generated
- How will data be generated
- What metadata are needed

2. Ethics, legal and security Issues



- How will ethical issues be handled
- How are the data accessed
- Are there copyright issues
- Are there sensitive data
- What about intellectual property rights

3. Data Storage and Preservation



- How are the data stored?
- Are there back up systems
- How are data safely preserved

4. Data Sharing and reuse



- How and where will the data be shared?
- How are sensitive data protected
- How can data be accessed

<https://www.uzh.ch/blog/hbz/2018/11/15/data-management-plan-in-a-nutshell/?lang=en>

2. Contexte Science Ouverte en pleine (r)évolution

- L'ANR a lancé un appel à projets "science ouverte" en 2019

Appel Flash *Science ouverte: pratiques de recherche et données ouvertes*

Les chiffres

- Ouverture de l'appel : 28/03/2019
- Clôture de l'appel : 27/05/2019
- 100 propositions reçues
- Budget alloué : 1,5 M€ budget initial + 0,8 M€ abondement MESRI
- Taux de sélection 25%

 [Lien vers page SO-DRIIHM](#)

- Page web dédiée [aux projets lauréats](#)
- Une étude des 100 propositions soumises à l'appel Flash
Disponible sur le site [Ouvrir la science](#)

→ Le projet
SO-DRIIHM
fait partie des
25 lauréats

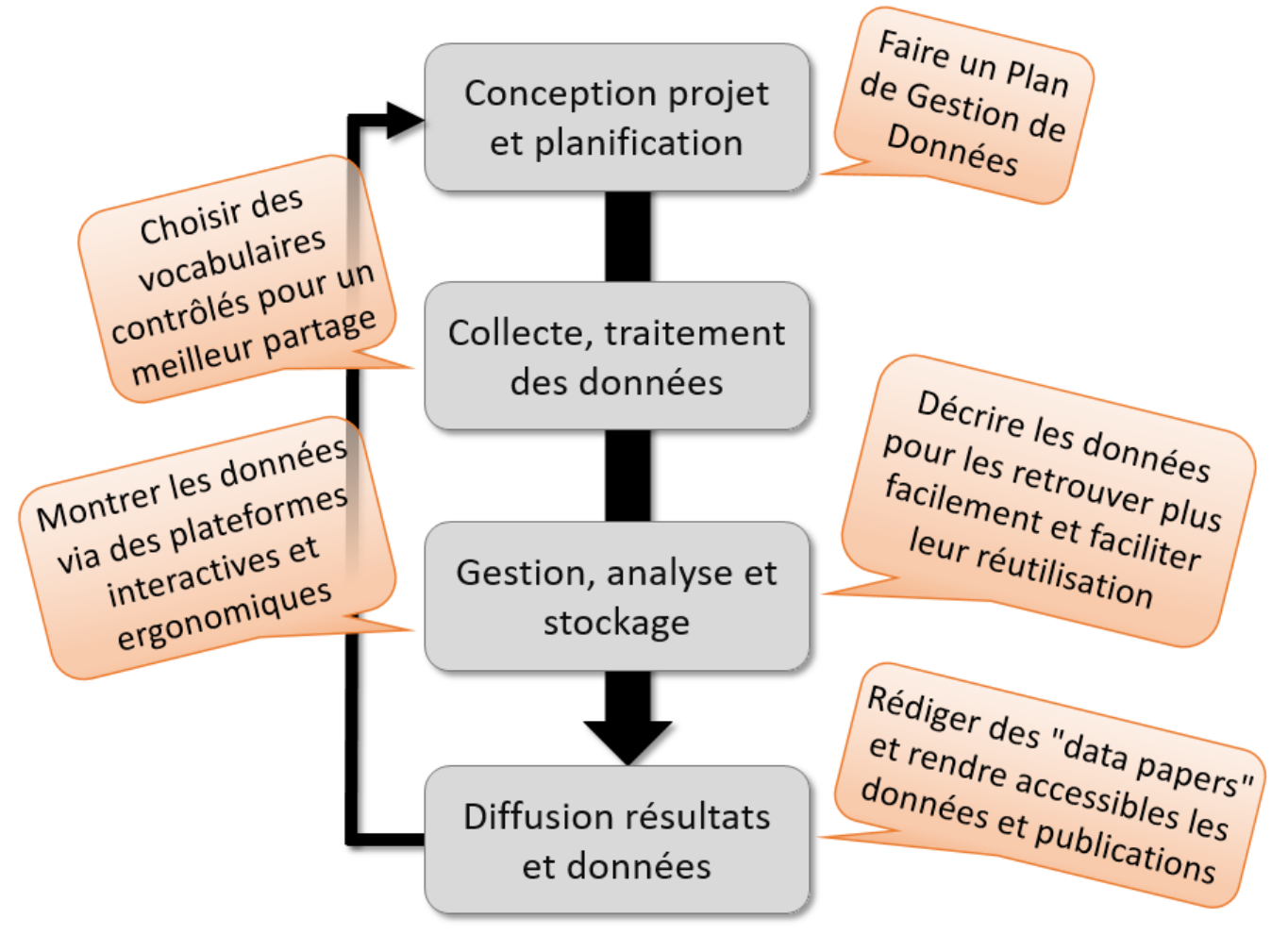
3. Le projet ANR SO-DRIIHM 2020-2023

2 grands objectifs :

- 1. Promotion de la science ouverte :** montrer les bénéfices du partage et de l'ouverture des données à la communauté scientifique DRIIHM
- 2. Co-construction d'une e-infrastructure** ergonomique et interopérable, intégrant les outils existants et accompagnant les chercheurs dans la "FAIRisation" de leurs données

"As open as possible, as closed as necessary"
(H2020 EU Commission)

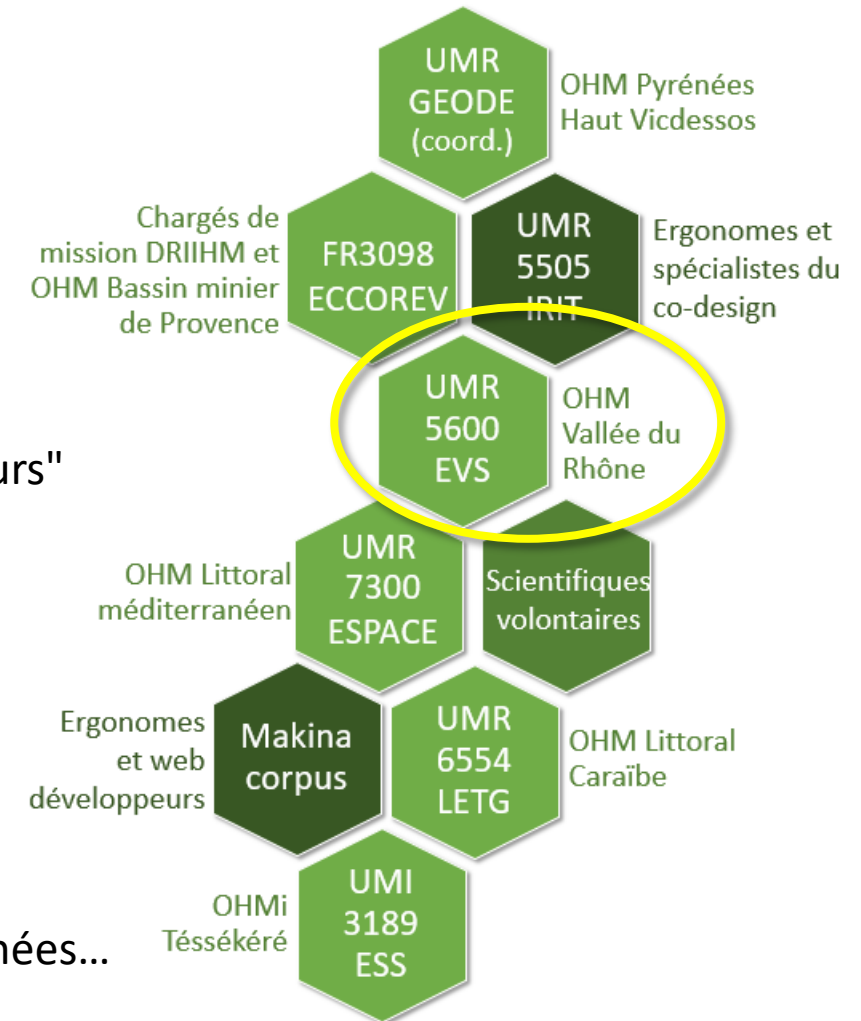
Agir tout au long du cycle de vie des projets de recherche



3. Le projet ANR SO-DRIIHM 2020-2023

● En bref - QQQQCCP

- **Quoi ?** Données de la recherche (à financement public à $\geq 50\%$)
- **Qui ?** Cible : 13 OHM du LabEx DRIIHM
 - Maîtrise d'ouvrage : membres du groupe data-DRIIHM + IRIT Toulouse (ergonomie, conception e-infrastructure)
 - Maîtrise d'œuvre : membres du groupe data-DRIIHM + CDD IE (12 mois) + Makina corpus (ergonomie, développ. web)
 - Acteurs : scientifiques producteurs et utilisateurs des données + "diffuseurs"
- **Où ?** Multi-sites
- **Quand ?** Octobre 2020 - mars 2023
- **Comment ?** Cf. plan d'action
- **Combien ?** 98 k€
- **Pourquoi ?** Mieux partager et ouvrir les données, permettre croisements données...



3. Le projet ANR SO-DRIIHM 2020-2023

- **Les grandes lignes du plan d'action**

- ✓ **Campagne de sensibilisation** à la politique de science ouverte et aux pratiques de science ouverte : PGD, rendre les données FAIR, déposer dans entrepôt de données, rédiger un data paper, etc.
- ✓ Tester la mise en oeuvre des bonnes pratiques de science ouverte à travers des “**projets pilotes**” :
 - **Trafic** (Fondation de France - OHM Littoral Caraïbe)
 - **PastServ** (APR 2020 - OHM HV)
 - **EvolSédimDeltaRhône** (APR 2020 - OHM VR)
 - Base de données **RhôneEco** (OHM VR)
 - **GéOyapock** (APR 2020 - OHM Oyapock)
 - Accompagnement des **post-doctorants du DRIIHM**
- ✓ **Développer l'infrastructure** par l'ajout de modules spécifiques itérativement, en fonction des priorités données par la Direction du DRIIHM et des OHM, et des besoins de la communauté
- ✓ **Faire tester les modules** du prototype d'infrastructure par des **chercheurs des OHM volontaires**
- ✓ **Campagne de formation** à l'utilisation des différents modules de l'infrastructure DRIIHM



Accompagnement dans la « FAIRisation » des données

3. Le projet ANR SO-DRIHM 2020-2023

● Accompagnement des projets pilotes tout au long du cycle de vie des données

5) Optimiser la réutilisation des données diffusées

→ réfléchir en amont sur le choix des variables, la standardisation des données et métadonnées (cf. 1, 2, 3)

4) Définir **quelles données diffuser**

+ Trouver un **entrepôt de données** où les diffuser/"stocker durablement"
+ Identifier d'autre(s) portail(s) pour diffuser les métadonnées (lisibilité ↗)
+ Définir **licence** accès et utilisation
+ Explorer piste d'un **data paper**



1) Créer un **Plan de gestion de données**

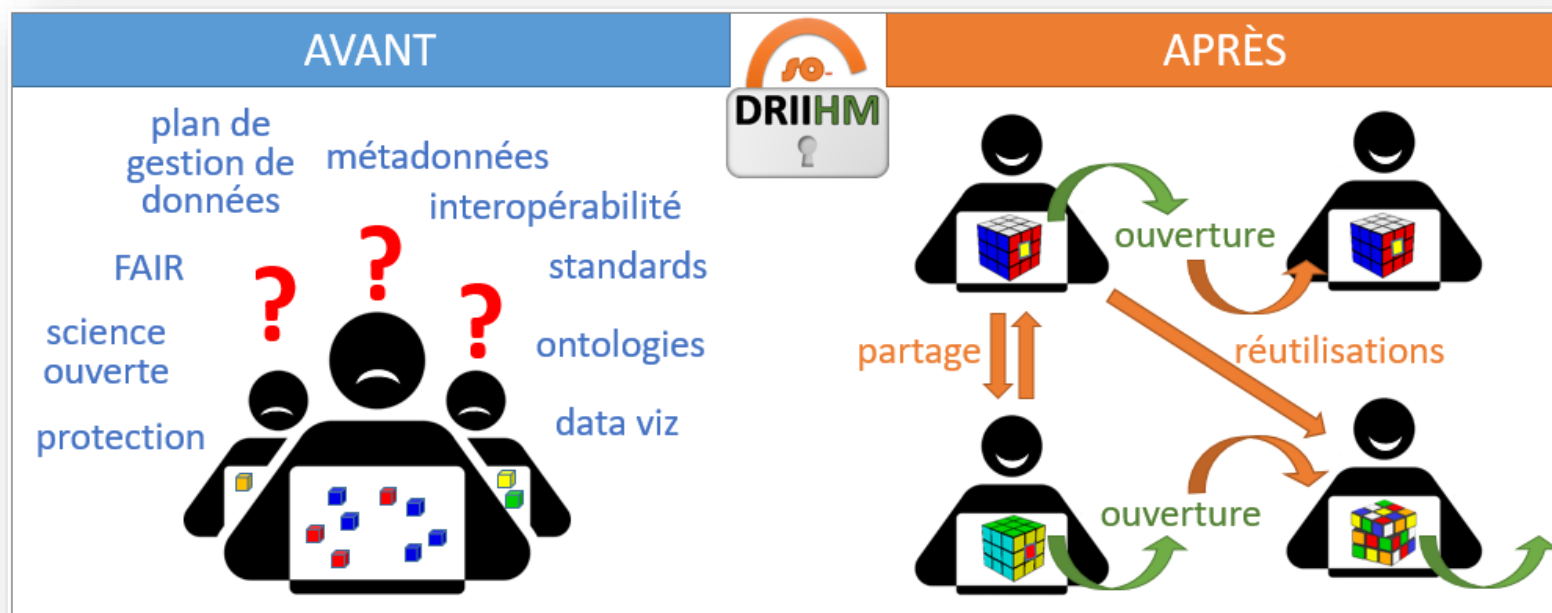
2) Réfléchir à l'**optimisation de la collecte** des données brutes, à leur **stockage**

3) Identifier des méthodes facilitant **nettoyage** et **traitements** des données brutes
+ Réfléchir à **standardisation** des données pour améliorer leur future réutilisation
+ Identifier un **standard de métadonnées**
+ Identifier où stocker/sauvegarder les scripts de traitement et les données traitées

D'après Research data lifecycle – UK Data Service
<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

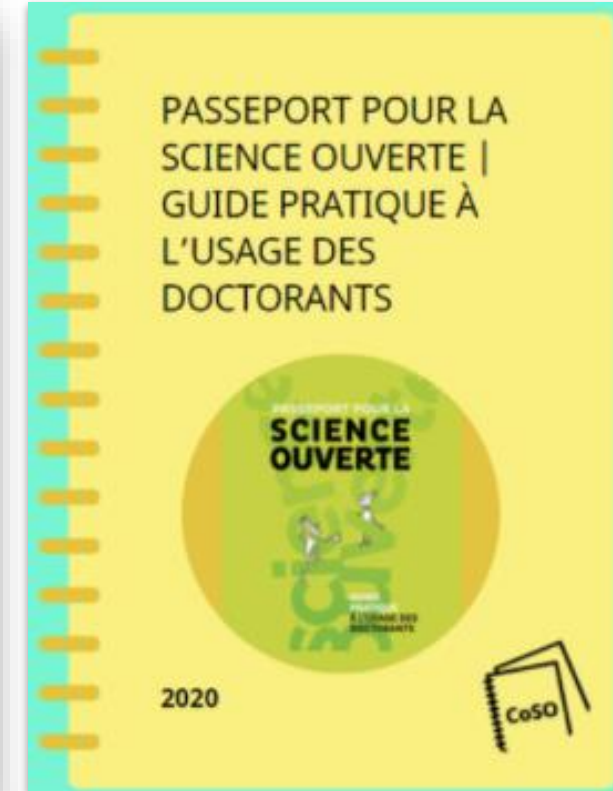
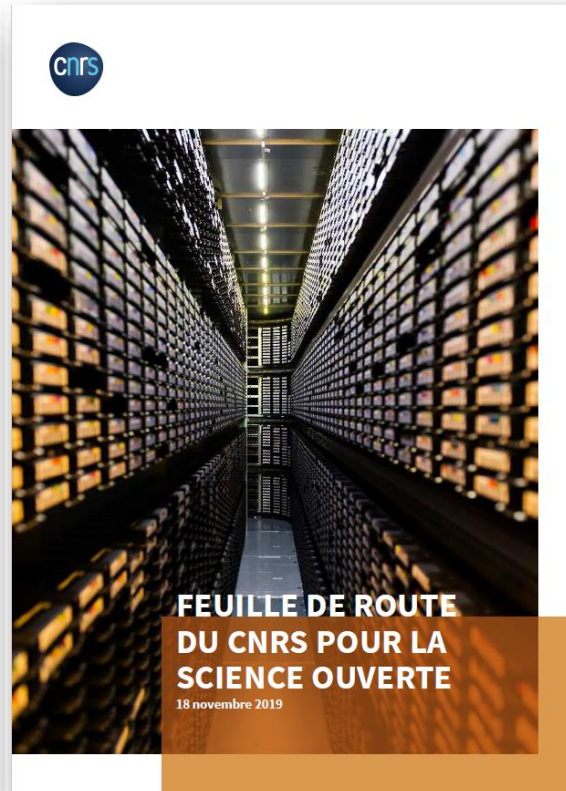
Conclusion

- Constat : dispersion productions scientifiques + relatif faible partage des données des OHM
- Un pas vers le partage et l'ouverture des données, mais nécessité d'accompagnement
- Contexte science ouverte (inter)national en plein boom
- Projet ANR SO-DRIIHM 2020-2023
 - Promotion de la science ouverte, formation aux bonnes pratiques : PGD, données "FAIR", ...
 - Co-construction d'un nouveau portail "aiguilleur du ciel" facilitant l'accès aux données des OHM et à leur "FAIRisation" → documenter ses données (métadonnées), c'est les rendre visibles !



Conclusion

Quelques documents de référence pour en savoir plus...



Nous contacter : datadriihm-groupe@services.cnrs.fr