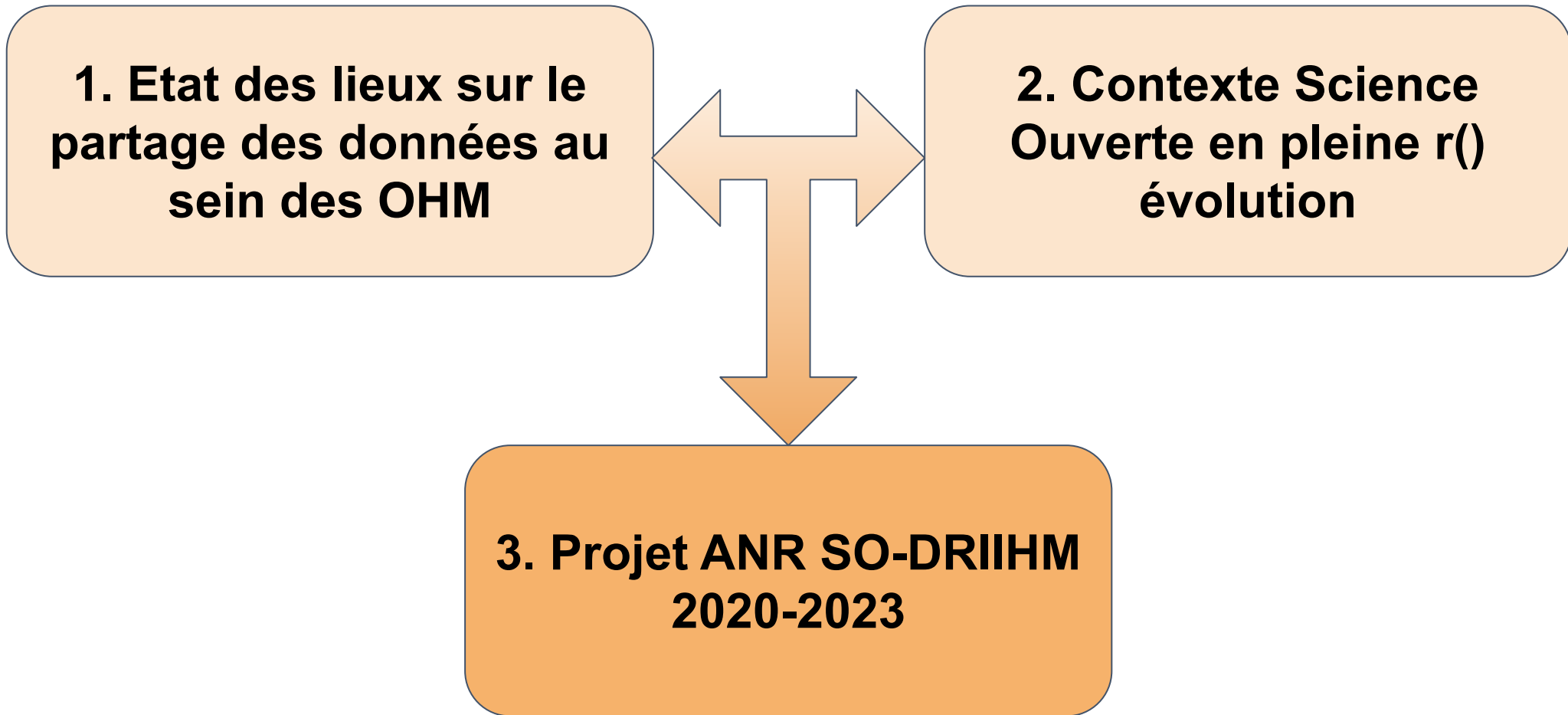


# Science ouverte et OHM :

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



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

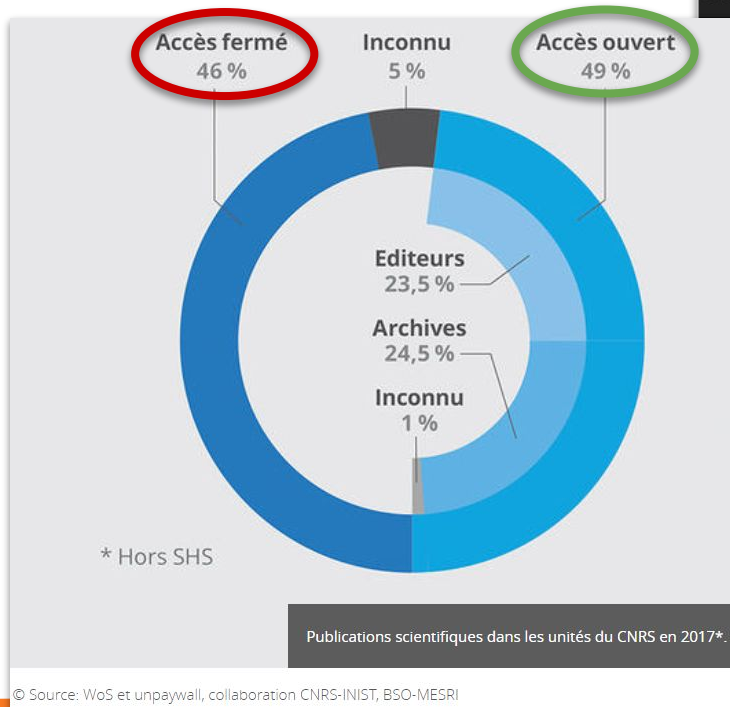


# Etat des lieux sur le partage des données au sein des OHM

## ● Constat 1 : dispersion et faible accessibilité des publications scientifiques des OHM

- Référencement dans archive ouverte HAL devenu obligatoire va améliorer la situation 😊
- Pas toujours de citation du LabEx DRIIHM (cf. charte OHM)
- Rappel des [consignes de dépôt dans HAL](#)

[Collection HAL OHM-BMP](#)




CCSD HAL Episciences.org Sciencesconf.org Support

LabEx DRIIHM OHM Bassin minier de Provence

### OHM Bassin Minier de Provence

Page d'accueil  
 Consultation des dépôts  
 Recherche

Recherche par documents

Filtres  
 Tous x

67 résultats  
 enregistrer la recherche

**TYPE DE DOCUMENT**

- Article dans une revue (28)
- Communication dans un congrès (20)
- Poster (10)
- Pré-publication,
- Document de travail (5)
- Rapport (2)
- Chapitre d'ouvrage (1)
- Ouvrage (y compris édition critique et traduction) (1)

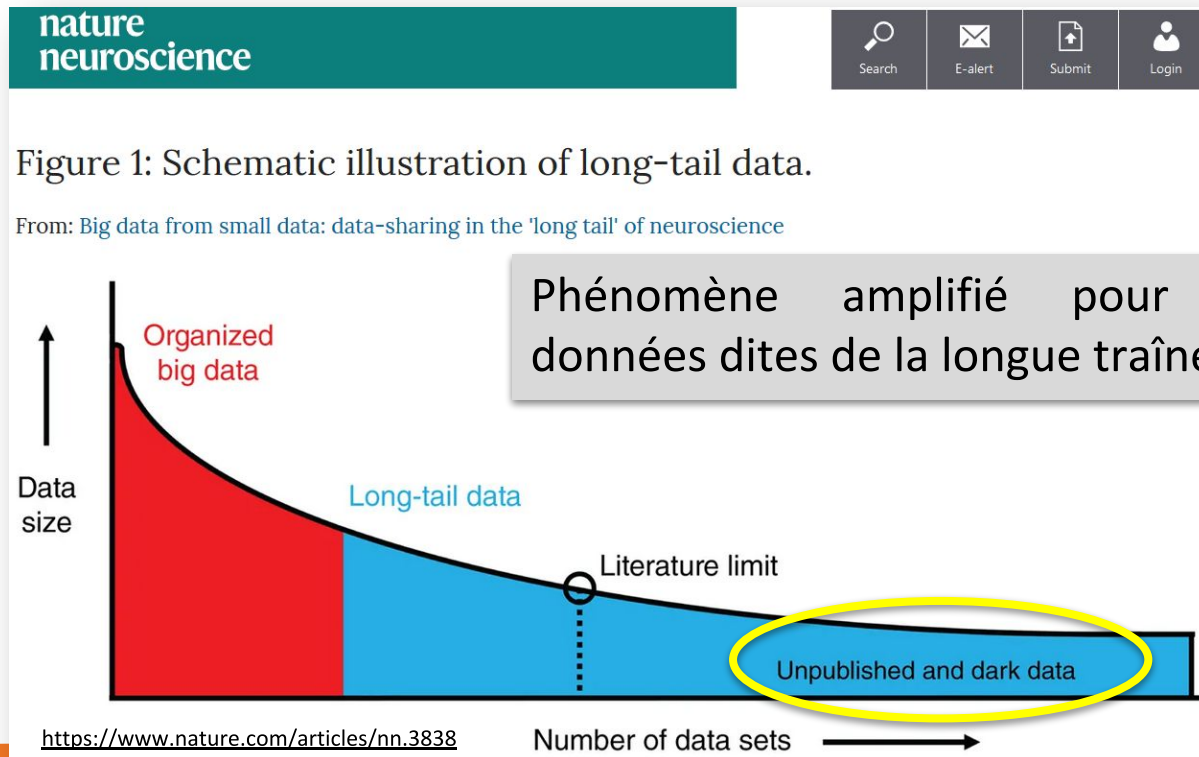
hal-02953717v1 Article dans une revue  
 Y. Noack. **Les Observatoires Hommes-Milieus, des outils de recherche interdisciplinaire au service du débat public. L'exemple de l'OHM Bassin Minier de Provence**  
*Mines et carrières, revue de l'industrie minière*, 2020, pp.33-37

hal-02953021v1 Article dans une revue  
 Hadrien Fanton, Evelyne Franquet, Maxime Logez, Nicolas Kaldonski. **Pomphorhynchus laevis manipulates Gammarus pulex behaviour despite salt pollution**  
*Freshwater Biology, Wiley*, 2020, 65 (10), pp.1718 - 1725. (10.1111/fwb.13573)

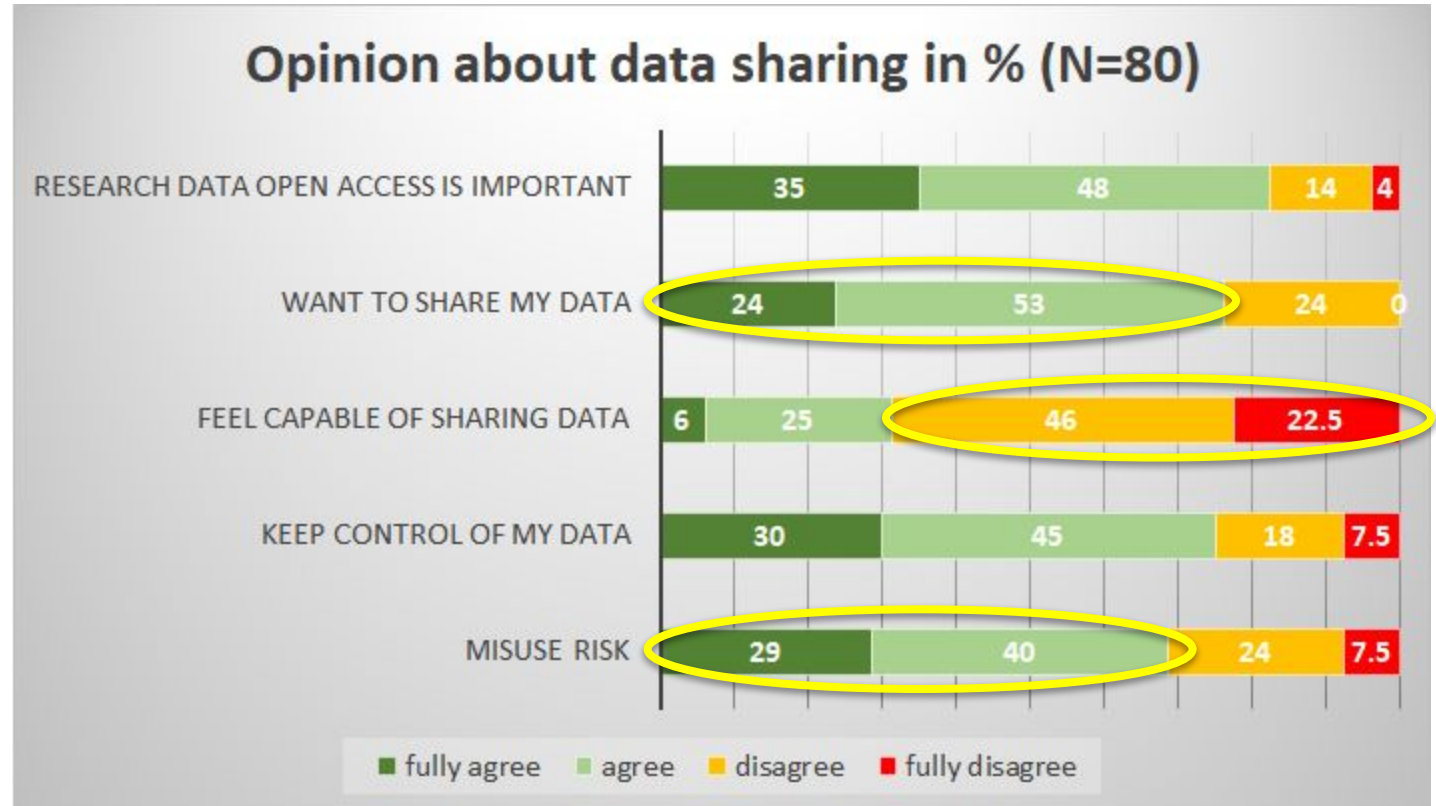
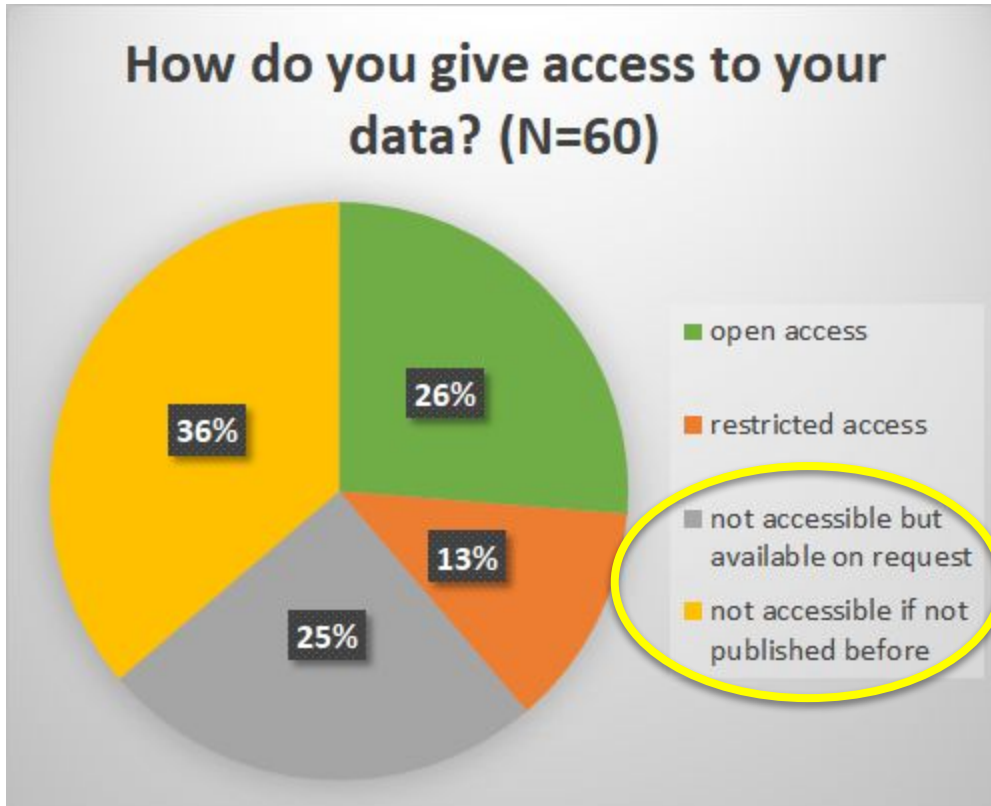
© Source: WoS et unpaywall, collaboration CNRS-INIST, BSO-MESRI

- **Constat 2 : dispersion, méconnaissance voire perte des données des OHM**

- données publiées “encapsulées” dans les PDF ou dans les “supplementary data”
- données publiées dans divers **entrepôts de données +/- accessibles** ou connus
- 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)
- données **peu accessibles via les géocatalogues** des OHM (dont [cat-BMP](#)) et du [géo-DRIIHM](#) central



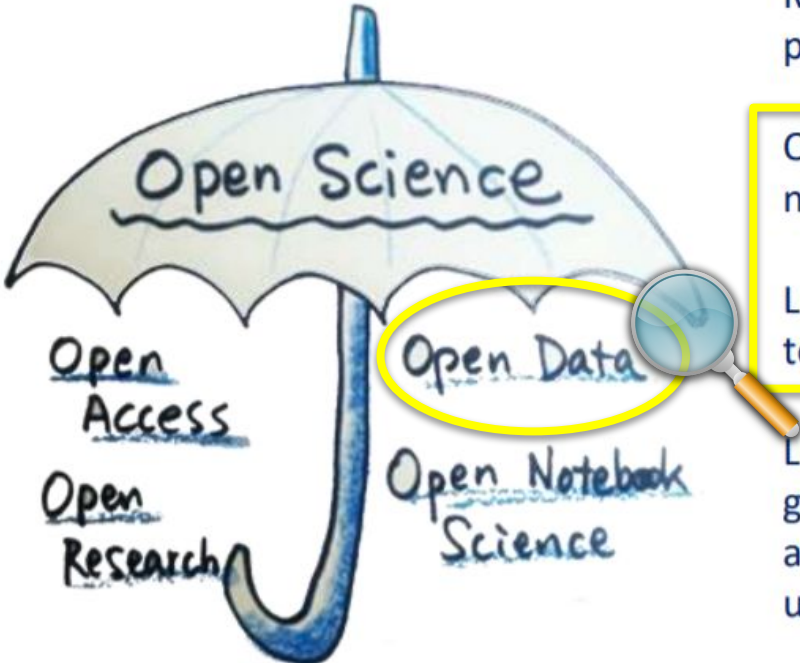
- **Enquête 2018** sur l'analyse du degré d'acceptabilité de la création d'une plateforme de partage des jeux de données pour les OHM



→ 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

## Science ouverte : un contexte 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...

 Depuis 80 ans, nos connaissances bâtissent de nouveaux mondes.

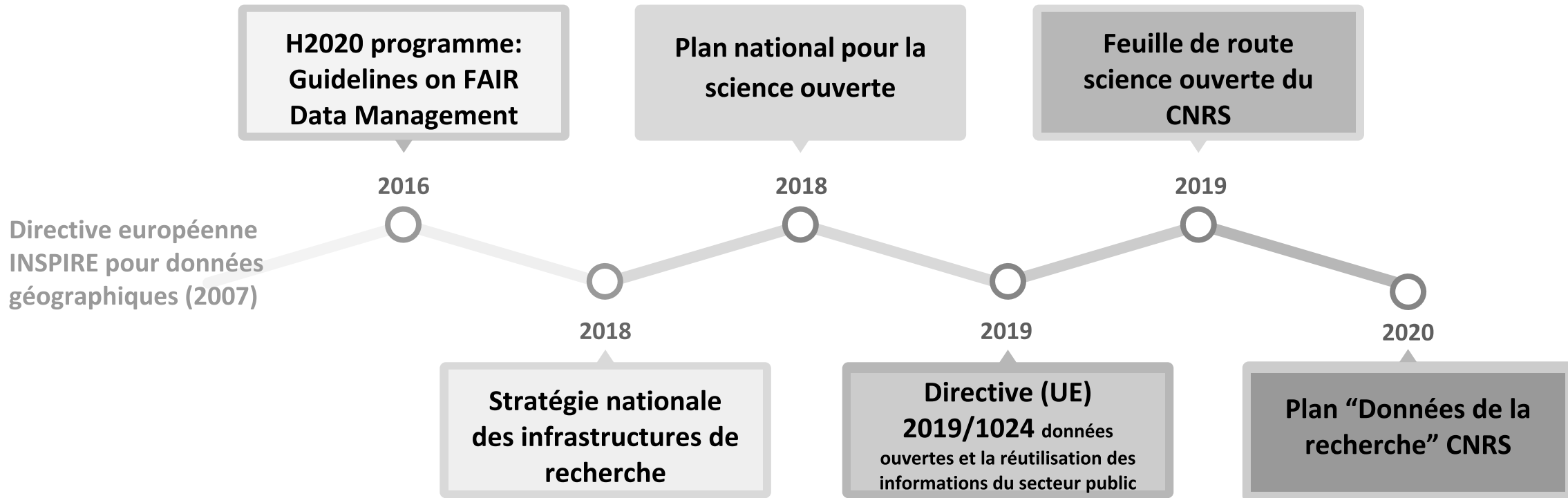
[https://inso2019.sciencesconf.org/data/pages/Lancement\\_Feuille\\_de\\_Route\\_SO\\_CNRS\\_AP\\_20191118.pdf](https://inso2019.sciencesconf.org/data/pages/Lancement_Feuille_de_Route_SO_CNRS_AP_20191118.pdf)

P 2

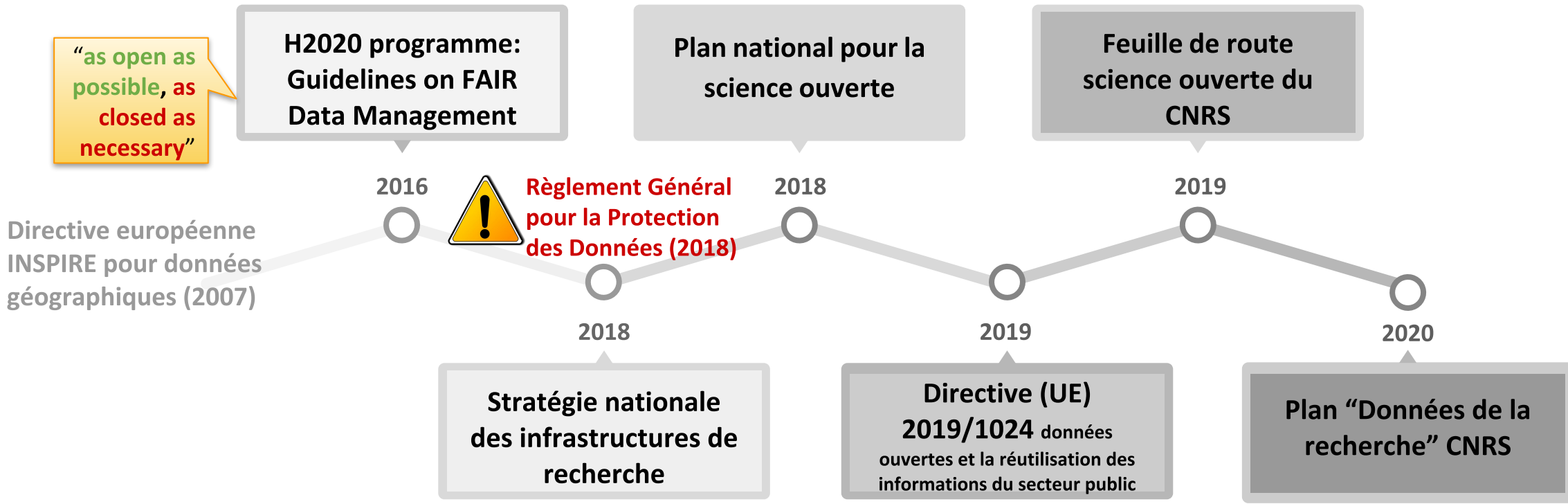
Intervention d'Antoine Petit, PDG du CNRS, aux Journées nationales de la Science Ouverte 2019

“La science ouverte, une révolution nécessaire”

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



- Principe “**ouvrir les données autant que possible, les fermer autant que nécessaire**” repris dans les différents textes





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

Obligation à utiliser **Plan de gestion de données** + **principes FAIR**

“Rendre obligatoire la **diffusion ouverte** des données de recherche issues de programmes financés sur fonds publics (sauf exceptions)”

“Développer une **culture de la gestion/partage** des données chez tous les acteurs du cycle de vie de la donnée basée sur la mise en œuvre des **principes FAIR**”

**H2020 programme: Guidelines on FAIR Data Management**

**Plan national pour la science ouverte**

**Feuille de route science ouverte du CNRS**

“faire **évoluer** les **pratiques** et les mentalités tout autant que **développer des outils** pour la **gestion, le partage, la préservation** à long terme et la diffusion des données de recherche en conformité avec les **principes FAIR.**”

Directive européenne INSPIRE pour données géographiques (2007)

**99 infrastructures** produisent, manipulent, traitent et/ou échangent des données

**Stratégie nationale des infrastructures de recherche**

**Directive (UE) 2019/1024** données ouvertes et la réutilisation des informations du secteur public

**Plan “Données de la recherche” CNRS**



“veiller à ce que la **planification** de la gestion des données deviennent une **pratique scientifique universelle** et soutenir la **diffusion** de données de la recherche qui soient traçables, accessibles, interopérables et réutilisables (**principes FAIR**)”

## ● Principes FAIR : augmenter le potentiel des données

Exemple fictif pour comprendre la “FAIRisation” d’un jeu de données  
(ici seules les métadonnées sont directement accessibles)

### 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 ? ...**
- 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**
- **Accès restreint aux données, sur demande (cf. contacts) mais accès ouvert aux métadonnées**
- **Licence d’utilisation CC BY-NC-ND 2.0**

- **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, d'un système d'information, d'une structure, d'une 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>

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

→ Le projet SO-DRIIHM fait partie des 25 lauréats

JC. Raynal pour  
l'OHM BMP

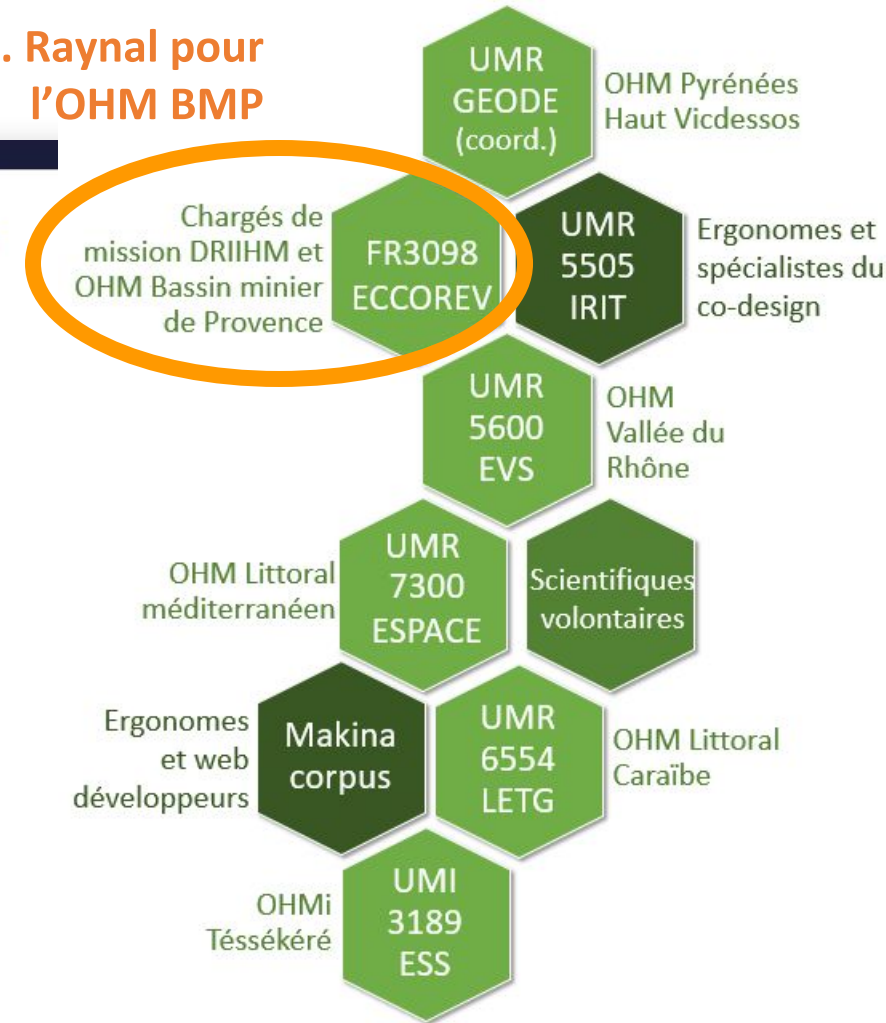
## Appel Flash Science ouverte: pratiques de recherche et données ouvertes

### 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 sur site](#)

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



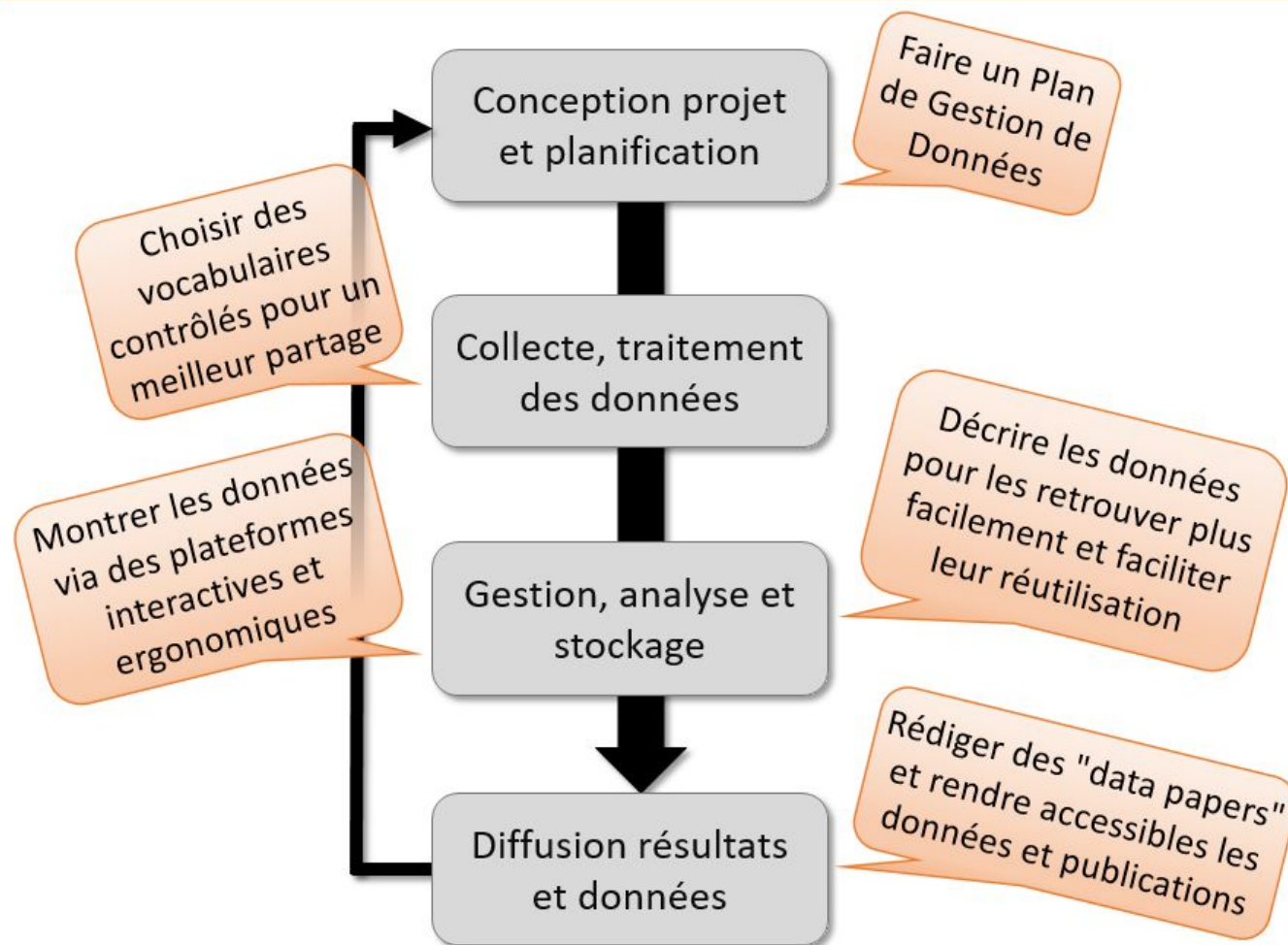
## 2 grands objectifs :

**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

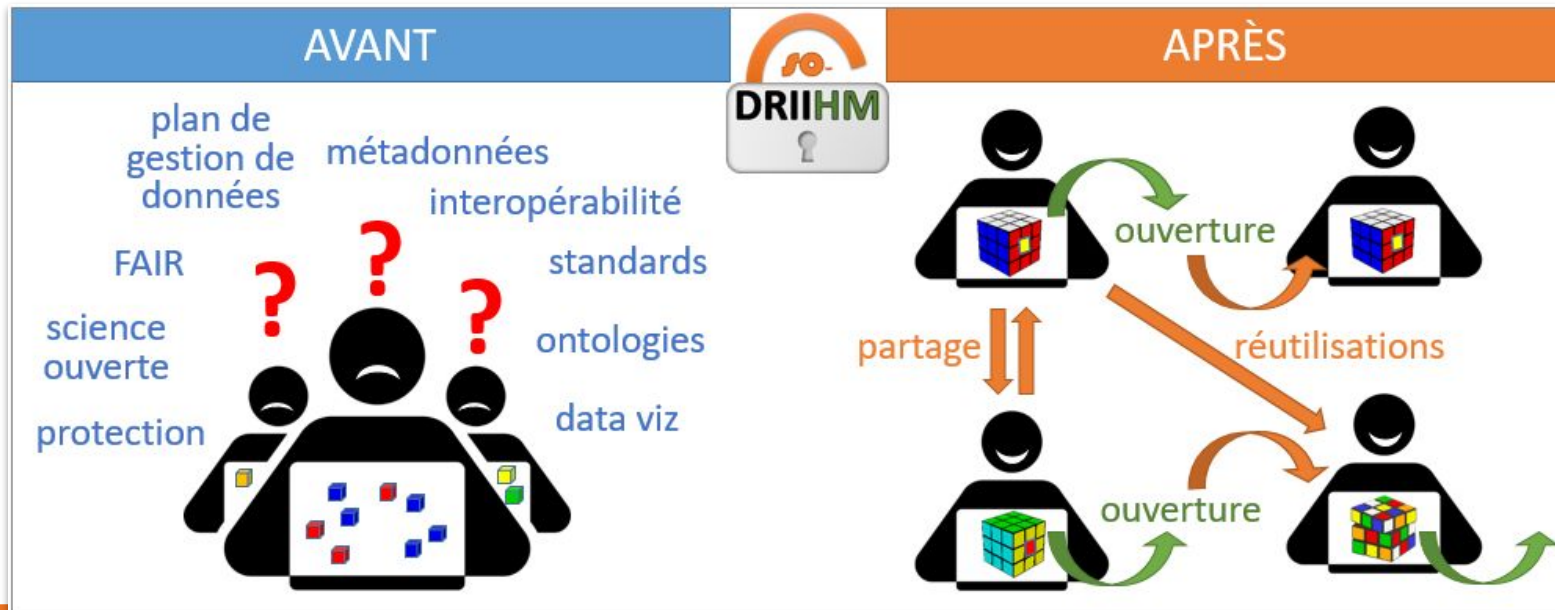


# Les grandes lignes du plan d'action de SO-DRIIHM

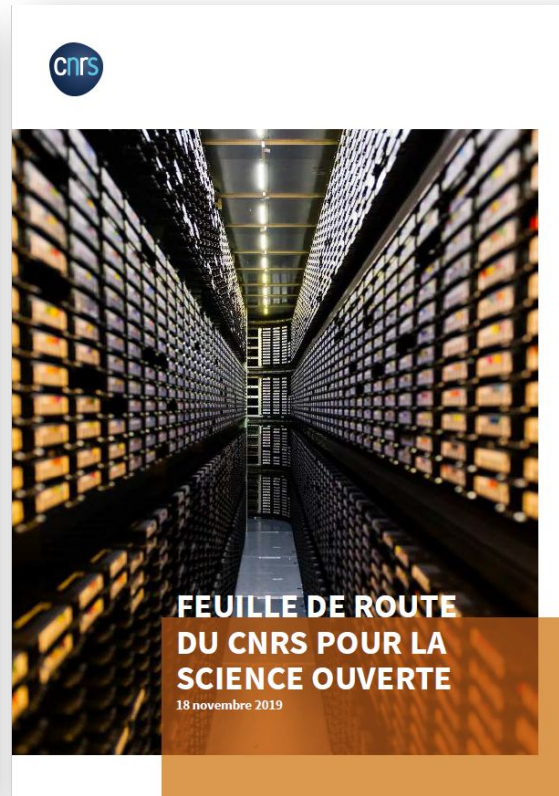
- **Campagne de sensibilisation** à la politique de science ouverte et aux pratiques de science ouverte : PGD, rendre les données FAIR, entrepôts de données, rédiger un data paper, etc.
- Tester la mise en oeuvre des bonnes pratiques de science ouverte à travers des “**projets pilotes**” :
  - **Trafic** (Fondation de France - OHM Littoral Caraïbe)
  - **PastServ** (APR 2020 - OHM HV)
  - **EvoSédimDeltaRhône** (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

# 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 !



## Quelques documents de référence pour en savoir plus...



Nous contacter : [datadrihm-groupe@services.cnrs.fr](mailto:datadrihm-groupe@services.cnrs.fr)