

# Open Science



Focus sur les entrepôts de données



Amorce sur le dépôt de logiciel

# UNESCO Recommendation on Open Science” (2021)

La science ouverte est

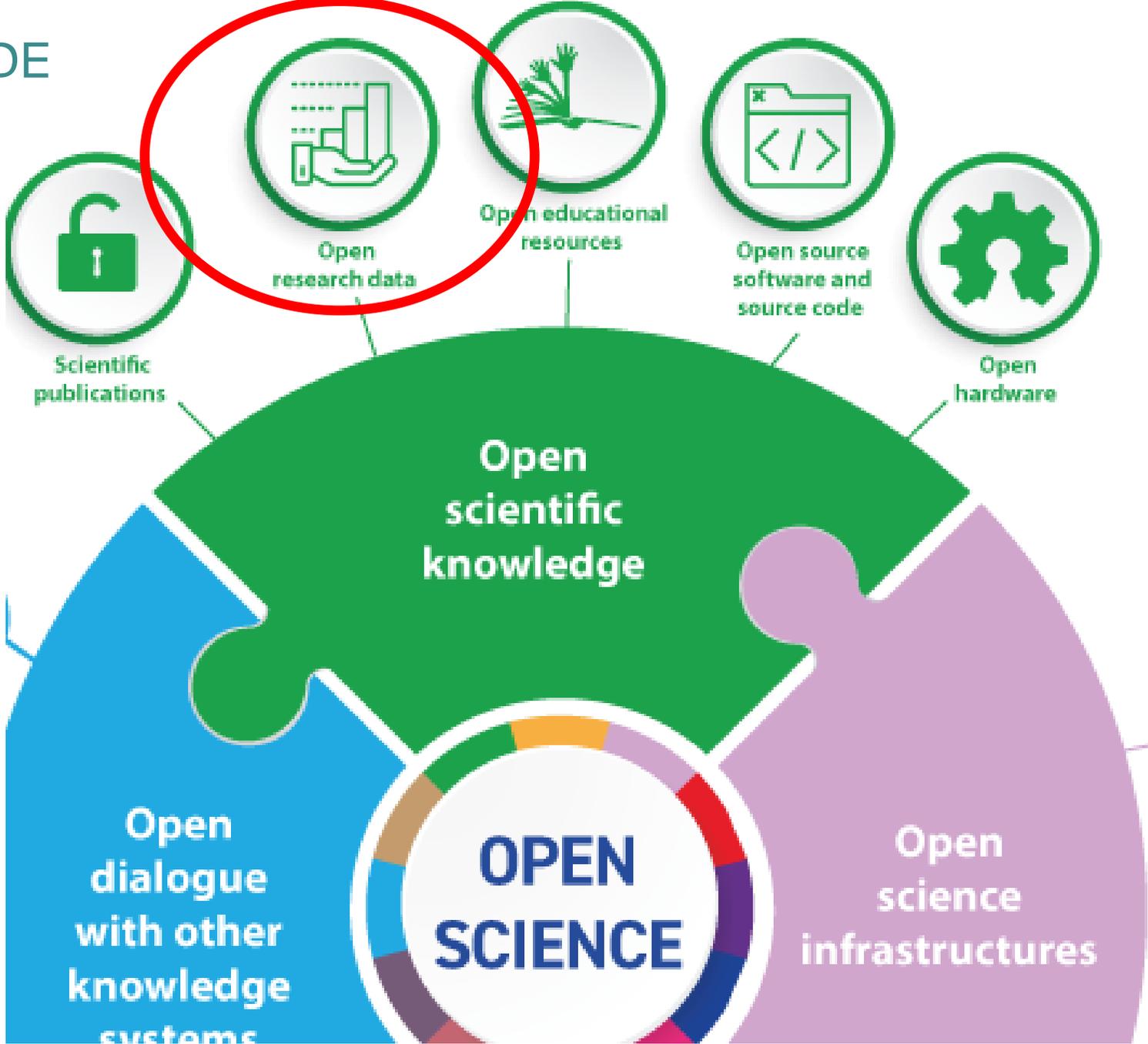
## un ensemble de principes et de pratiques

qui visent à rendre la recherche scientifique dans tous les domaines accessible à tous, au bénéfice des scientifiques et de la société dans son ensemble.

*En promouvant une science plus accessible, inclusive et transparente, **la science ouverte favorise le droit de chacun à participer au progrès scientifique et à ses bienfaits**, comme le stipule l’article 27.1 de la Déclaration universelle des droits de l’homme.*



# OUVERTURE DES DONNEES DE LA RECHERCHE



# DEFINITION DONNEES DE LA RECHERCHE

Les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires **pour valider des résultats de recherche**.

*Définition proposée par l'OCDE dans son rapport « [Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics de 2007](#) ».*

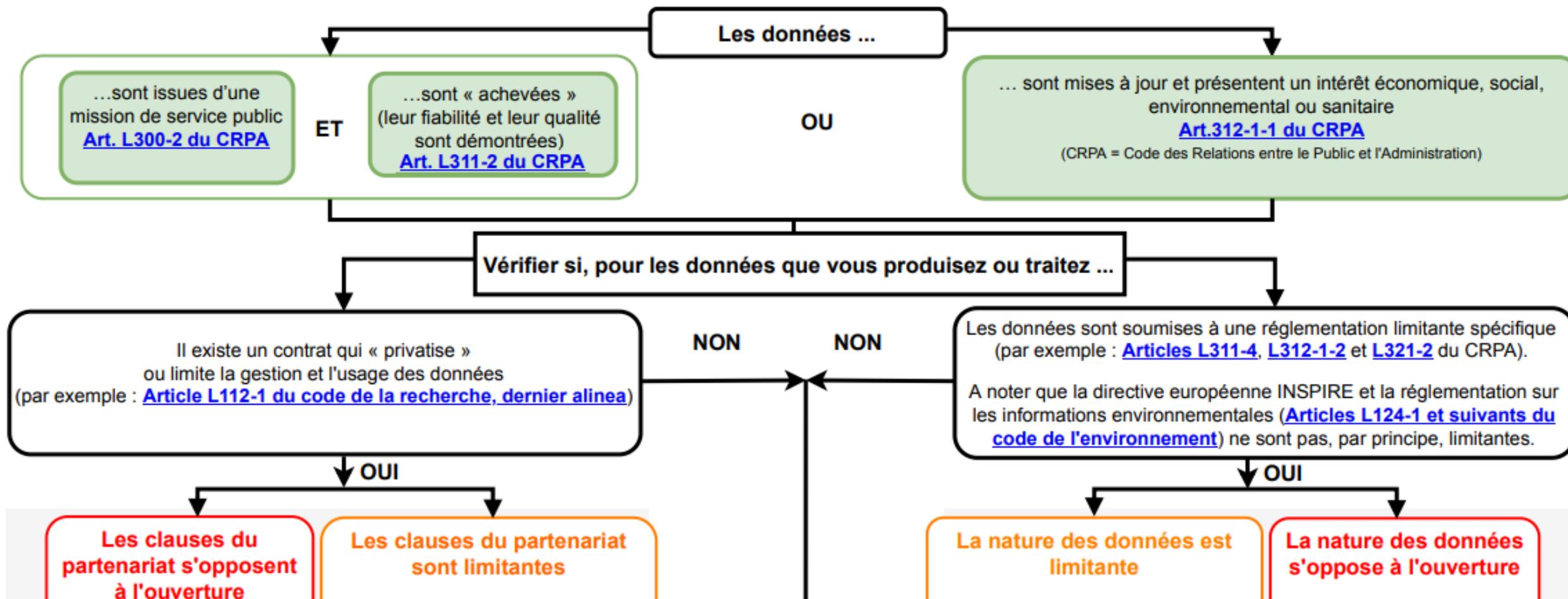
Le périmètre des données de la recherche est circonscrit par ce texte à la notion d'éléments probants, nécessaires **à la validation du processus de recherche**.

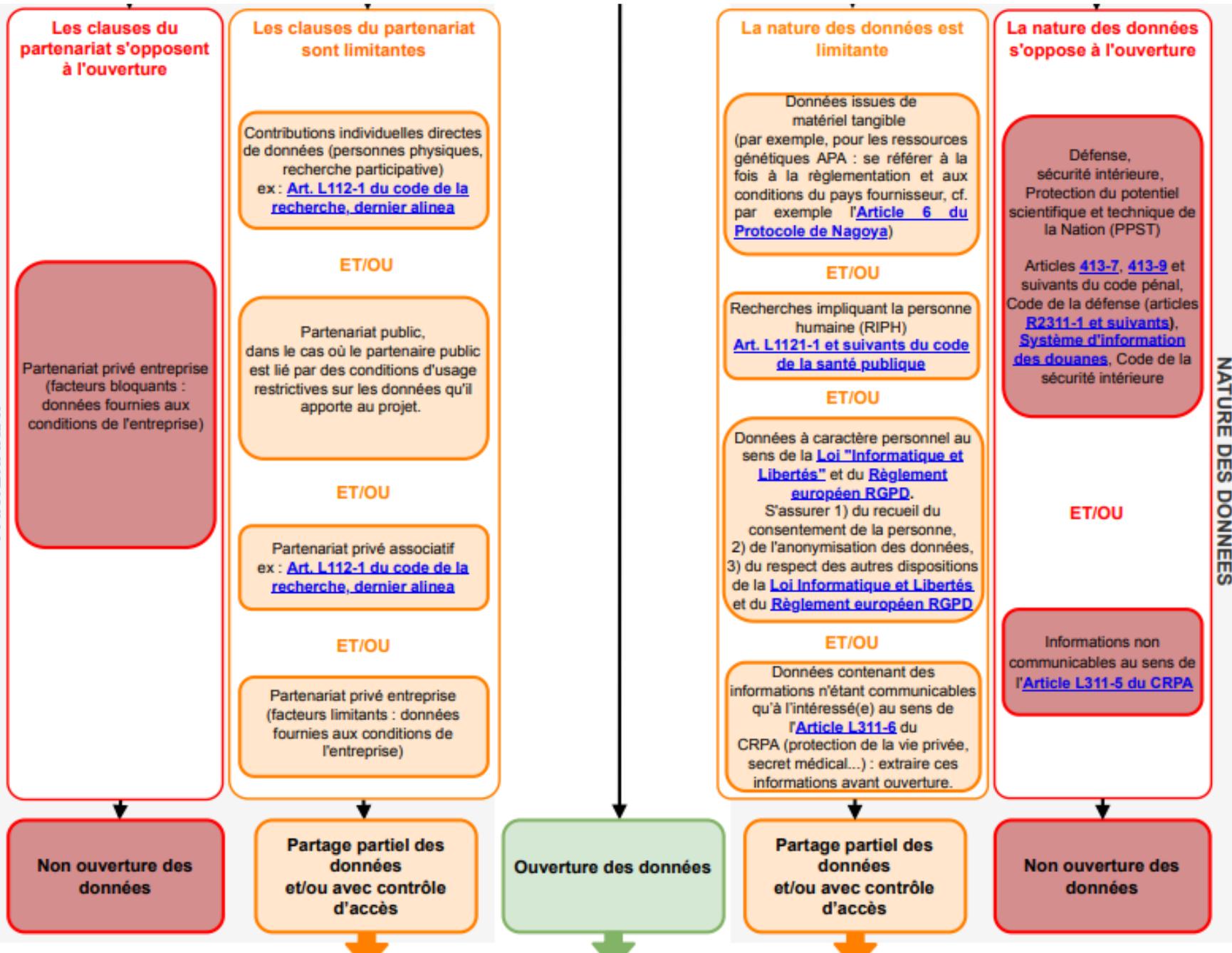
→ Cela permet de déterminer ce qui devra ou non être conservé de manière pérenne.

Cette définition **exclut de son périmètre les cahiers de laboratoires** alors que ceux-ci sont explicitement mentionnés dans les modèles de plans de gestion de données mis à disposition par les financeurs.

En effet, les cahiers de laboratoire sont des éléments essentiels pour documenter le contexte de production des données. De surcroît, les cahiers de laboratoire font partie des documents à archiver de manière pérenne, d'après le référentiel des archives de la recherche.

# OUVERTURE DES DONNÉES : ... AUSSI OUVERT QUE POSSIBLE ... AUSSI FERMÉ QUE NÉCESSAIRE





#### LES QUESTIONS À SE POSER EN COMPLÉMENT

- Stratégie scientifique ?
- Stratégie de partenariat ?
- Questions de déontologie (voire d'éthique) ?
- Impact possible négatif sur environnement et/ou dans le domaine socio-économique ?
- Délai pour exploiter pleinement sur le plan scientifique et publication des résultats ?

# \* Le cycle de vie des données

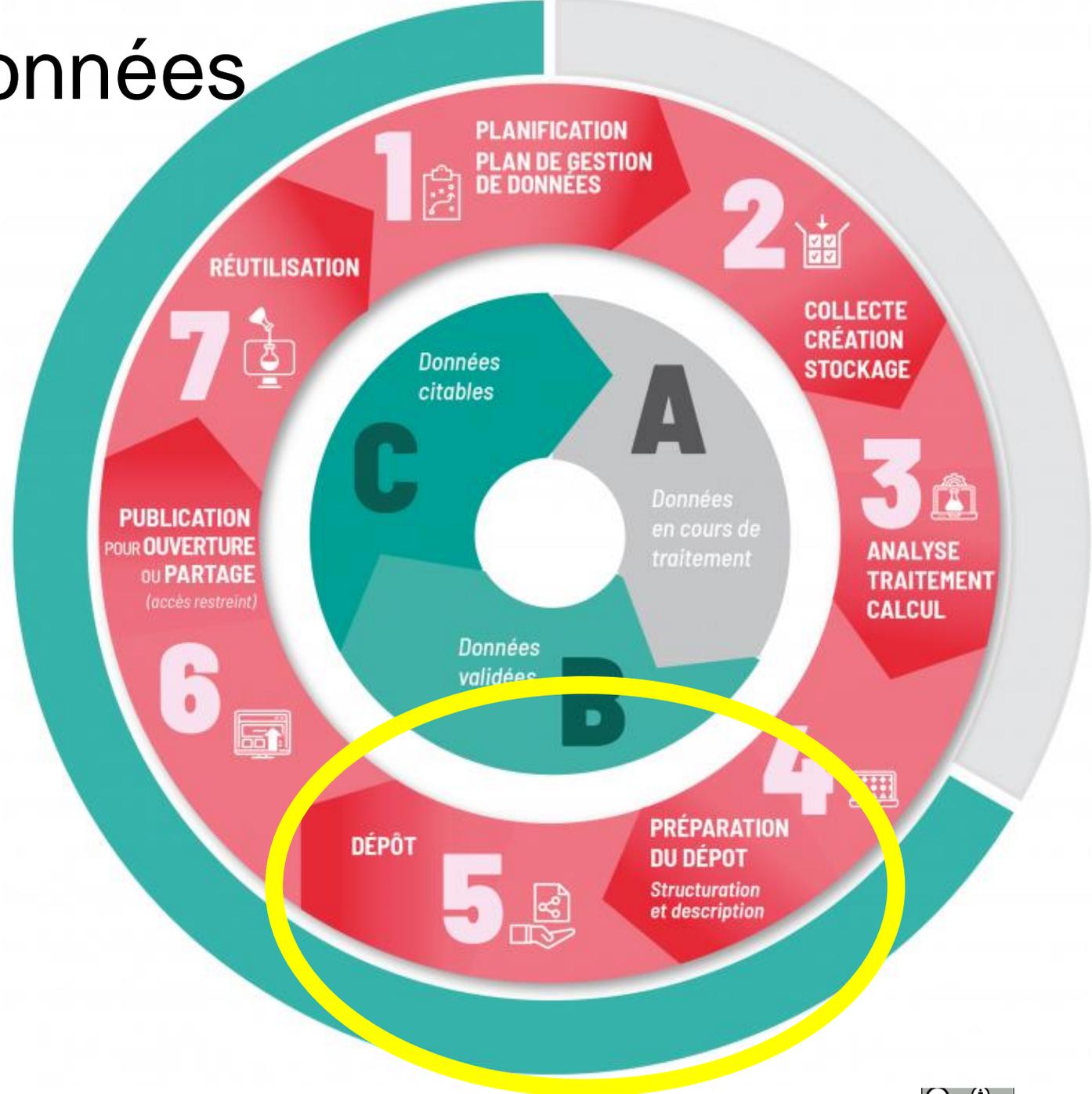
## Principe FAIR

**FAIR (Findable, Accessible, Interoperable, Reusable)**

Les principes FAIR visent à rendre les données

- Faciles à trouver,
- Accessibles
- Interopérables
- Réutilisables.

**Findable, Accessible, Interoperable, Reusable**



# \* Entrepôt de données : lequel choisir ?

Le choix d'un entrepôt de données sera guidé par

- les pratiques de votre communauté scientifique
- la politique de votre établissement et la disponibilité d'un entrepôt institutionnel
- les instructions aux auteurs de la revue dans laquelle vous soumettez votre article.

Plusieurs répertoires d'entrepôts de données permettent d'affiner la sélection en appliquant différents critères.

# \* Exemple d'entrepôts de données

Exemples d'entrepôts de données :

- Institutionnels (en France) : [Dataverse cirad](#), [Datapartage](#) (INRAe), [DataSuds](#) (IRD)

→ **Recherche.data.gouv**

- Pluridisciplinaires et internationaux : [Zenodo](#), [Dryad](#), [Figshare](#)

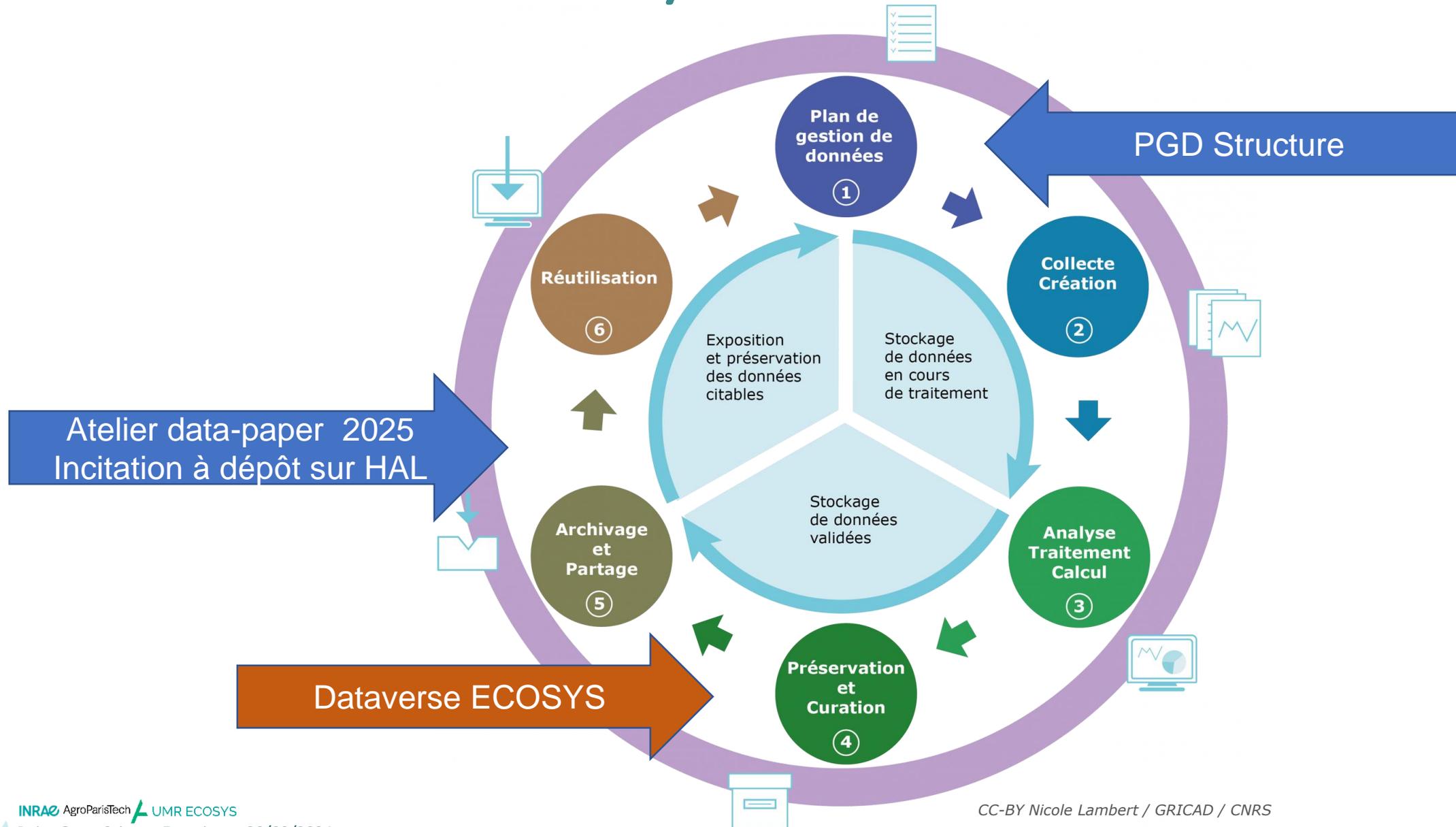
- Thématiques ou disciplinaires :

[GenBank](#) (séquences génétiques), [TRY](#) (caractères botaniques), [GBIF](#) (biodiversité), [Pangaea](#) (sciences de la terre et de l'environnement), [WormBase](#) (nématologie), [Movebank](#) (mobilité animale), [West African Vegetation](#), [DataFirst](#) (enquêtes socio-économiques en Afrique), [Protocols.io](#) (protocoles), etc.

- Liés à un éditeur :

[GigaDB](#) (Oxford Univ. Press), [Dataverse Ubiquity Press](#), [Dataverse Economics](#)

# Où en est-t-on à Ecosys ? <sup>1/2</sup>



# \* Démonstration dataverse Ecosys

<https://entrepot.recherche.data.gouv.fr/dataverse/ecosys>



UMR\_ECOSYS Dataverse Données de recherche unité écologie fonctionnelle et écotoxicologie des agroécosystèmes / services écosystémiques  
(www.inrae.fr)

Recherche Data Gouv > Data INRAE > Experimental - Observation - Simulation Dataverse >

Contact Partager Modifier

Chercher dans cette collection...



Recherche avancée

+ Ajouter des données

- Collections (1)
- Jeux de données (35)
- Fichiers (241)

Catégorie de collection  
Groupe de recherche (1)

Année de publication  
2024 (4)  
2023 (11)

1 à 10 de 36 résultats

Tri

Chemical characterization of volatile organic compounds emitted by animal manure

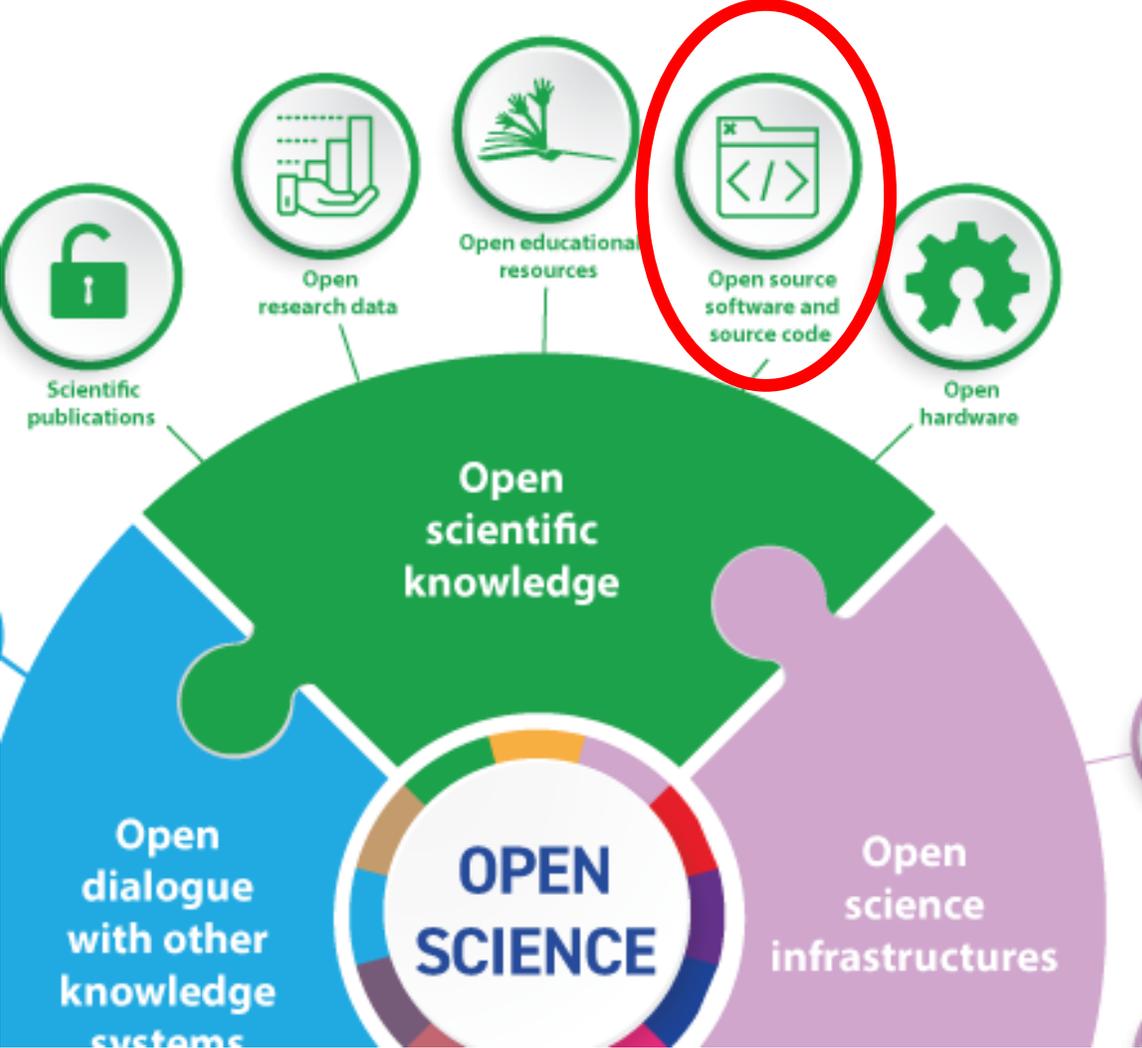
3 mai 2024



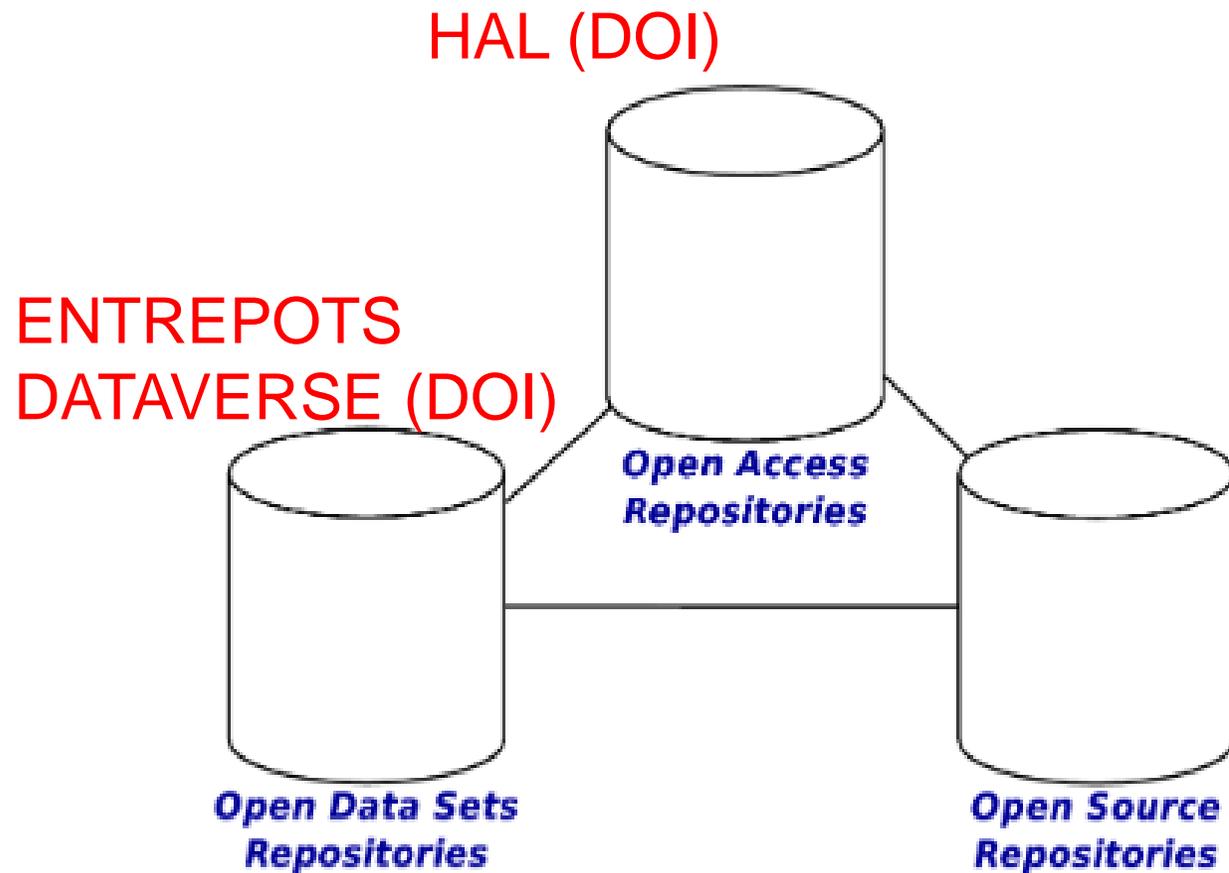
ciurar, raluca, 2024, "Chemical characterization of volatile organic compounds emitted by animal manure",  
<https://doi.org/10.57745/PGB2SY>, Recherche Data Gouv, V1

Experimental investigation involving four animal manure samples: cows, horses, sheep and goats. This data file includes a complete inventory of volatile organic compounds emitted from four animal manure samples and detected by Thermal Desorption-Gas Chromatography-Mass Spectromet...

# Amorce sur les pratiques pour l'ouverture des logiciels scientifiques/des code source



# \* Pour une reproductibilité des résultats de la recherche : les trois piliers



La préservation des logiciels est à la base de la **reproductibilité**. L'utilisation des logiciels intervient à toutes les étapes de la recherche, dans tous les domaines scientifiques, et se révèle essentielle à de multiples égards. Pour reproduire une expérience, il est indispensable de connaître avec exactitude la version du logiciel employé.

Software Heritage viendra garantir la **pérennité** et la **traçabilité** des logiciels sur lesquels s'appuient les travaux de recherche, permettant d'assurer la reproductibilité de leurs résultats.

[Software Heritage](#)

(**SWHID**) (lien avec forge)



Software Heritage

# \* Quelques chiffres

Les 

## 25 %

C'est l'augmentation moyenne du nombre de citations observée lorsqu'un jeu de données déposé dans un entrepôt est lié à une publication.

*Colavizza & al., 2020.*

Un réel enjeu

## 10 milliards d'euros

C'est la perte minimale que représentent les données non-partagées par l'ensemble de la recherche européenne (liée au temps passé à re-produire des données, aux coûts de stockage, de licences, de double financements...).

*Cost of not having FAIR research data, 2018*

## - INRAE

Formation e-learning :

<https://elearning.formation-permanente.inrae.fr/course/view.php?id=668>

Site dédié à la science ouverte :

<https://science-ouverte.inrae.fr/fr>

- Autres sites d'intérêt (MESR, etc...)

<https://www.canal-u.tv/chaines/ouvrirlascience/passeport-pour-la-science-ouverte/serie-passeport-initiation-a-la-science>

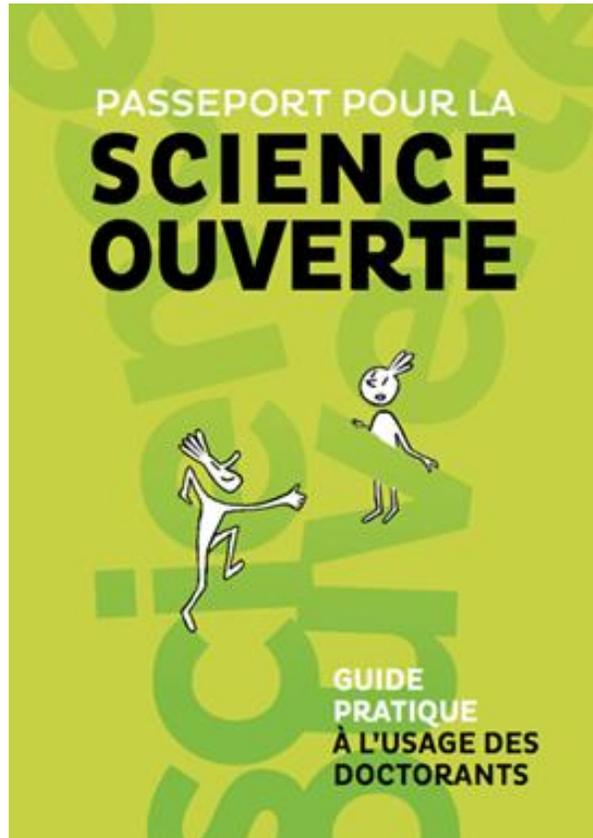
<https://www.ouvrirlascience.fr/serie-passeport-initiation-a-la-science-ouverte/>

- Collection « Passeport pour la science ouverte » avec des livrets thématiques

- Livrets thématiques "Codes et logiciels" (<https://www.ouvrirlascience.fr/science-ouverte-codes-et-logiciels/>) et

"Entrez dans le débat" (<https://www.ouvrirlascience.fr/science-ouverte-entrez-dans-le-debat/>).

(<https://www.ouvrirlascience.fr/former-a-la-science-ouverte-tout-au-long-de-la-these/>).



# RDO

## Référents Données Opérationnels

*Des collègues missionnés pour vous assister dans la gestion des données de la recherche*

Merci pour votre attention

Place aux questions

