



## Principes FAIR et cycle de vie des données

**Date de dernière mise à jour : novembre 2025**

→ Institut de l'information scientifique et technique  
DVDR | Service Formation-DoRANum

Ce webinaire a pour objectif de vous expliquer à quoi correspondent les principes FAIR. Ces derniers s'appliquent tout au long du cycle de vie des données de recherche. Celui-ci sera détaillé en expliquant à chaque étape les bonnes pratiques à mettre en place pour respecter les principes FAIR.

# Sommaire

01 Définition des données de recherche

02 Principes FAIR

03 Cycle de vie des données de recherche

Étape 1 : planification (PDG / DMP)

Étape 2 : création, collecte et description des données de recherche

Étape 3 : traitement et analyse, stockage sécurisé durant le projet

Étape 4 : partage, diffusion des données, dépôt dans un entrepôt

Étape 5 : archivage pérenne

Étape 6 : réutilisation et valorisation des données

04 À retenir



01

## Définition des données de recherche

## Définition des données de 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 les résultats** de recherche. »

(Définition l'OCDE)

Si vous voulez en savoir plus : [What is data?](#)



4

Plusieurs définitions des données de la recherche existent.

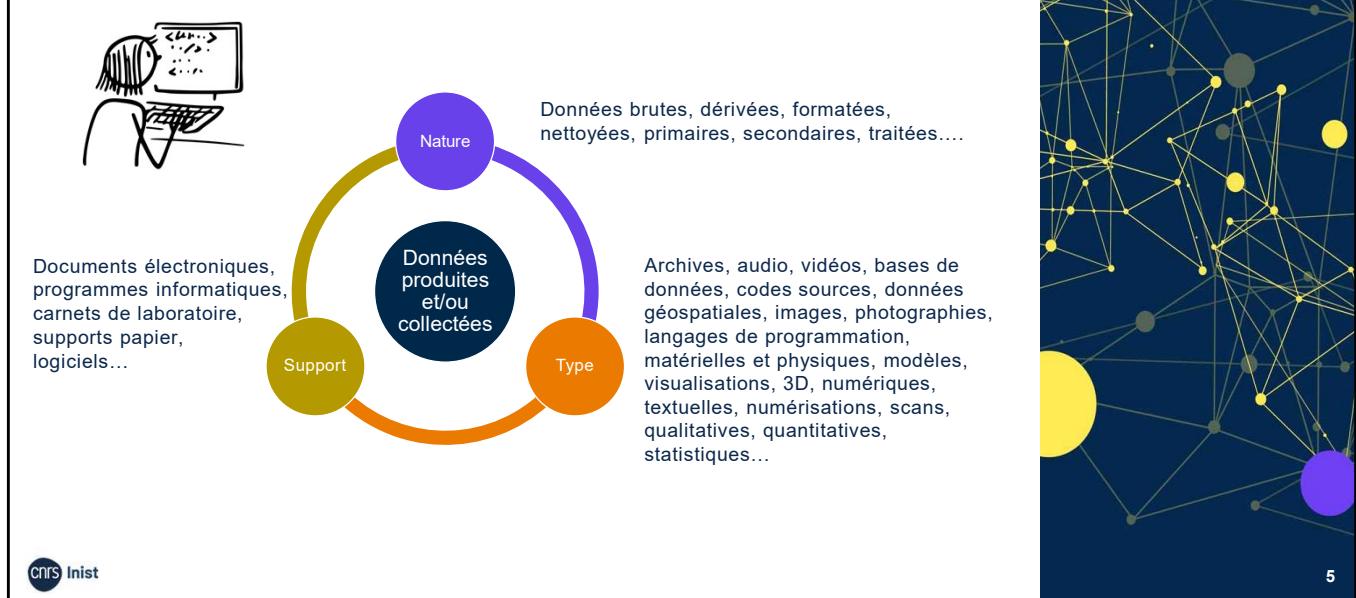
« Les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales pour la recherche scientifique** et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider les résultats de recherche.** » (OCDE, *Organisation de Coopération et de Développement Économiques. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. 2007.*

<https://doi.org/10.1787/9789264034020-en-fr>)

On peut également citer la définition issue de l'ARDC - Australian Research Data Commons (traduction Inist-CNRS) :

« Le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé [...]. Les données peuvent être numériques, descriptives, visuelles ou tactiles. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou support [...]. » (ARDC, *Australian Research Data Commons. What Is Research Data. 2019*)

# Définition des données de recherche



Selon leur contexte de création (capture ou production), leur exploitation, leur analyse et les traitements qu'elles subissent, les données de recherche peuvent être :

- de différentes **natures** : données brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées....
- de tous **types** : audio, vidéos, images, bases de données, codes sources, données géospatiales, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques...
- contenues dans divers **supports** : documents électroniques, programmes informatiques, carnets de laboratoire, supports papier, logiciels...

## Sources :

- Rivet Alain, Bachèlerie Marie-Laure, Denis-Meyere Auriane, Tisserand Delphine. *Traçabilité des activités de recherche et gestion des connaissances. Guide pratique de mise en place.* 2018. <https://qualite-en-recherche.cnrs.fr/gt/tracabilite-des-activites-de-recherche/>
- UNIL, Université de Lausanne. *Les données de recherche.* <https://www.unil.ch/unil/fr/home/menuinst/recherche/open-science/open-research-data/gerer-les-donnees-de-recherche/les-donnees-de-recherche.html>
- Image : Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

**Ressource** : DoRANum. L'origine et la description des données de la recherche.

<https://doi.org/10.13143/E9ZH-W908>

# Exemples de données de recherche



Photo d'une peinture pariétale représentant une scène de chasse

Code Provenance accessoires	Code Provenance	Pays d'origine	Nom de la provenance possible dans l'échantillon	Autre nom de provenance	Conditions géographiques (m)	Altitude (m)	P (mm)	substrat	Poids de l'échantillon (g)	% graines vives immobiles + transportées dans l'eau	Présence					
											01-4800- 01-01	01-4800- 01-02	01-4800- 02-01	01-4800- 02-02	01-4800- 02-03	
6	7	6578	Espagne	Andalousie	361200; 370000; 375000	650	1000	grès	653.6	4.8	x	x	x	x	x	
9	6579	Espagne	Messica	371300; 375000	710	Campos	654.2	3.2	-	x	x	x	x	x	x	
10	6580	Espagne	Madrid	371300; 375000	450	Campillo	568.2	12.1	-	x	x	x	x	x	x	
11	6581	Espagne	Silvesca / Chababdi	407140; 21734; 50-400	430	gravier	689.7	0.4	-	x	x	x	x	x	x	
12	7393	Italie	Ferragudo (PT)	372000; 375000	680	dépôts	714.3	12.7	-	x	x	x	x	x	x	
13	3913	Italie	Ferragudo (PT)	372000; 375000	680	grès	714.3	12.7	-	x	x	x	x	x	x	
14	6582	Italie	Monte Saccarello	371300; 375000	1000	grès/basalte	734.1	6.2	-	x	x	x	x	x	x	
15	18	6583	Liban	Bet Mousa / Bet Marouza	347200; 370000; 1300-1400	1000	grès/basalte	1341.7	8.2	-	x	x	x	x	x	x
16	6584	Liban	Qadisha	371300; 375000	1000	grès/basalte	714.3	12.7	-	x	x	x	x	x	x	
17	11	6585	Grèce	Sifnou / Chalakidi	371300; 375000	600-700	grès/basalte	784.3	7.8	-	x	x	x	x	x	x
18	6586	Grèce	Sifnou / Chalakidi	371300; 375000	600-700	grès/basalte	784.3	7.8	-	x	x	x	x	x	x	
19	13	6587	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
20	30	36558	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36559	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36560	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36561	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36562	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36563	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36564	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36565	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36566	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36567	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36568	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36569	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36570	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36571	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36572	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36573	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36574	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36575	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36576	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36577	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36578	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36579	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36580	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36581	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36582	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36583	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36584	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36585	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36586	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36587	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36588	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36589	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36590	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36591	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36592	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36593	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36594	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36595	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36596	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36597	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36598	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36599	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36600	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36601	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36602	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36603	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36604	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36605	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36606	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36607	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36608	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36609	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36610	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36611	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36612	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36613	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36614	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36615	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36616	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36617	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36618	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36619	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x	x	x
21	30	36620	Portugal	Norte do Tejo	40-410; 754-400	1296	gravier	912.4	4.8	-	x	x	x	x		

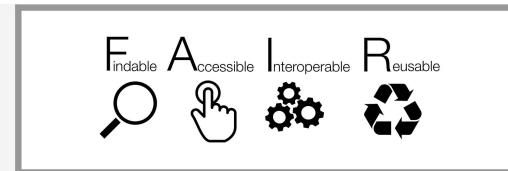
02

## Principles FAIR

# Principes FAIR

**4 principes** à respecter pour garantir une utilisation optimale des données de recherche et des métadonnées associées, à la fois **par les hommes et par les machines**

- **F** (Findable) = Facile à trouver
- **A** (Accessible) = Accessible
- **I** (Interoperable) = Interopérable
- **R** (Reusable) = Réutilisable



Principes admis par les différentes communautés scientifiques au niveau international, ainsi que par les financeurs (ex : Commission européenne, ANR, etc.)

**Applicables tout au long du cycle de vie des données**



8



Tout au long de cette présentation des petites bulles vous permettront de repérer à quel(s) principe(s) FAIR correspondent principalement les bonnes pratiques décrites.

**Source :**

*Image : Sangya Pundir [CC BY-SA 4.0]*

**Ressource :**

*DoRANum. Les principes FAIR. <https://doi.org/10.13143/Z7S6-ED26>*

## Principe F

**Le principe F** facilite la découverte des données par les humains et les systèmes informatiques :

- Catalogues ou entrepôts
- Identifiant pérennes
- Métadonnées riches



**Le principe F** signifie que vos données doivent être faciles à trouver par des humains et par des systèmes informatiques.

**Bonnes pratiques :**

- Déposer vos jeux de données dans un entrepôt de données
- Leur attribuer un identifiant pérenne
- Les décrire précisément à l'aide de métadonnées

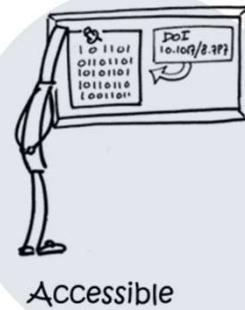
**Sources :**

- INRAe, *Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR.* <https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair>
- Image : Patrick Hochstenbach / Modifié par : DoRANum (CC-BY 4.0)

## Principe A

**Le principe A** encourage à stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint)

- Protocoles standardisés et ouverts
- Authentification et autorisation si besoin
- Accès permanent aux métadonnées



Accessible



10

**Le principe A** encourage à stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint).

### Bonnes pratiques :

- Protocoles de communication standards : les données doivent être rendues accessibles soit via une URL, soit via un téléchargement de fichier(s). L'idée est de ne pas avoir besoin de recourir à des outils spécialisés ou propriétaires pour pouvoir accéder aux données, exemples : Skype, Microsoft Exchange...
- Préférer le dépôt dans des entrepôts certifiés qui proposent un accès ouvert.
- Définir qui a le droit d'accéder aux (méta)données, par quelle procédure (ouverture totale ou demande d'accès par authentification ou demande d'autorisation)
- Les métadonnées restent accessibles même si les données ne le sont pas ou plus.

### Sources :

- INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. *Produire des données FAIR*. <https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair>
- Université Paris-Saclay. *Produire des données FAIR*. <https://www.universite-paris-saclay.fr/recherche/science-ouverte/produire-des-donnees-fair>
- Image : Patrick Hochstenbach / Modifié par : DoRANum (CC-BY 4.0)

## Principe I

**Le principe I** signifie que les données sont téléchargeables, utilisables, intelligibles et combinables avec d'autres données, par des humains et par des systèmes informatiques

- Formats ouverts
- Vocabulaires contrôlés (standards)
- Lien vers d'autres (méta)données



**Le principe I** signifie que les données sont téléchargeables, utilisables, intelligibles et combinables avec d'autres données, par des humains et des systèmes informatiques.

### Bonnes pratiques :

- Utiliser des formats ouverts et indépendants.
- Utiliser des vocabulaires standards, des thésaurus, des ontologies... Le but étant de « parler le même langage ».
- Contextualiser les données en indiquant les liens vers d'autres données et métadonnées ou vers des publications.

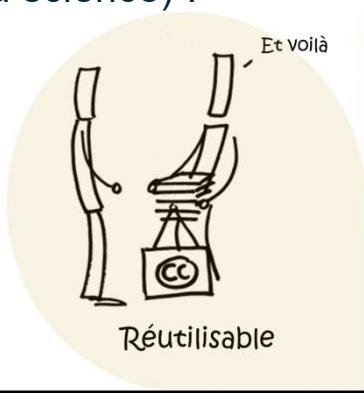
### Sources :

- INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR. <https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair>
- Image : Patrick Hochstenbach / Modifié par : DoRANum (CC-BY 4.0)

## Principe R

**Le principe R** encourage à rendre vos données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et transparence de la science) :

- Description riche
- Provenance
- Standards communautaires
- Licences



12

**Le principe R** encourage à rendre vos données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et transparence de la science). Les métadonnées et les données doivent être bien décrites afin de pouvoir être reproduites et/ou combinées dans différents contextes.

### Bonnes pratiques :

- Les données sont très précisément décrites avec le plus d'informations possibles (métadonnées).
- Leur provenance est renseignée.
- Elles répondent à des standards communautaires pertinents pour le domaine.
- Il est important de leur attribuer une licence d'utilisation.

### Sources :

- INRAe, *Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR.* <https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair>
- Image : Patrick Hochstenbach / Modifié par : DoRANum (CC-BY 4.0)

## Qu'apportent les Principes FAIR ?

### Pour le chercheur

- Mise en place de bonnes pratiques
- Gain de temps dans la gestion d'un projet de recherche
- Meilleure visibilité et citation
- Amélioration de l'intégrité scientifique
- Données plus accessibles
- Collaborations favorisées



cnrs Inist

13

Afin de mieux saisir à quoi servent les principes FAIR, voici une liste de ce qu'ils apportent concrètement au chercheur ainsi qu'à la communauté scientifique.

### Pour le chercheur

- Il adopte de bonnes pratiques de gestion.
- C'est aussi un gain de temps dans la gestion du projet de recherche.
- Le chercheur gagne en visibilité.
- L'intégrité scientifique est facilitée.
- Ses données sont plus accessibles.
- Cela favorise les collaborations.

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Qu'apportent les Principes FAIR ? Pour la communauté scientifique

- Accès facile à des données publiques et réutilisables
- Meilleure interopérabilité des données
- Gain de temps et d'argent : ne pas recréer des données déjà existantes
- Meilleure reproductibilité de la recherche



### Pour la communauté scientifique

- L'accès à des données publiques et réutilisables est facilité, ainsi que l'accès à des corpus utiles pour d'autres domaines.
- L'interopérabilité des données est meilleure.
- Il y a également un gain de temps et d'argent puisque des données déjà existantes ne sont pas recréées.
- Cela permet une meilleure reproductibilité de la recherche.

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

03

## Cycle de vie des données de recherche

# Cycle de vie des données de recherche

C'est « l'ensemble des étapes

- de gestion
- de conservation
- de diffusion des données de recherche associées aux activités de recherche. »

Source : CIRAD



16

« Le cycle de vie des données de la recherche est l'ensemble des étapes de gestion, conservation, diffusion et réutilisation des données scientifiques, associées aux activités de recherche. »

*Deboin Marie-Claude. Découvrir de nouveaux métiers liés aux données de la recherche. CIRAD. 5 p. 5 octobre 2018. <https://doi.org/10.18167/coopist/0061>*

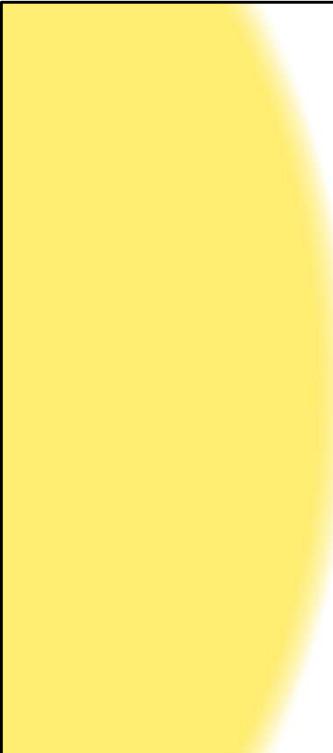
Nous allons détailler les 6 étapes dans notre présentation.

Source de l'image : d'après *Research data lifecycle – UK Data Service*.

Ressource :

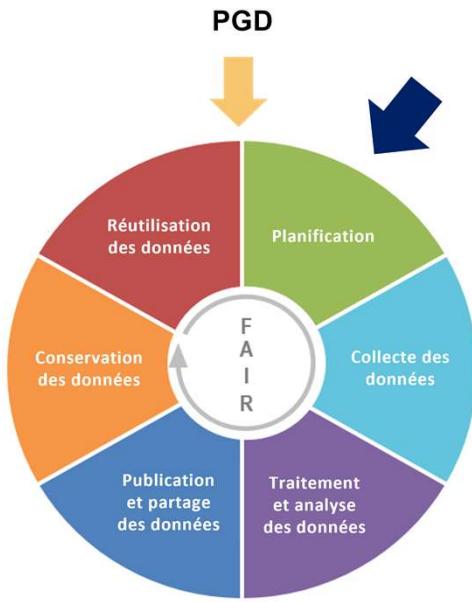
*DoRANum. Le cycle de vie des données de recherche.*

<https://doi.org/10.13143/GZJ2-J593>



## **Étape 1 : Planification (PGD/DMP)**

## Place dans le cycle de vie des données



CNRS Inist

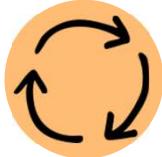
18

Il s'agit de la première étape du cycle de vie des données. C'est dès cette étape que commence la rédaction du Plan de Gestion de Données (PGD).

**Source de l'image :** d'après *Research data lifecycle – UK Data Service*.

## Le PGD qu'est-ce que c'est ?

PGD (Plan de Gestion des Données) = DMP (Data Management Plan)



Aide à organiser et anticiper toutes les étapes du cycle de vie



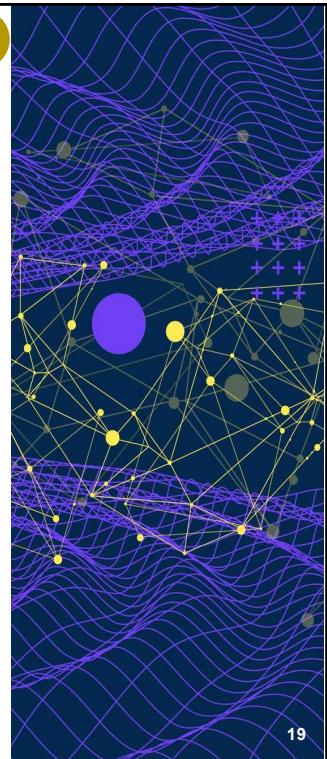
Détermine les actions à mettre en place pour une bonne gestion des données



Est un document évolutif



Aide à rendre les données FAIR

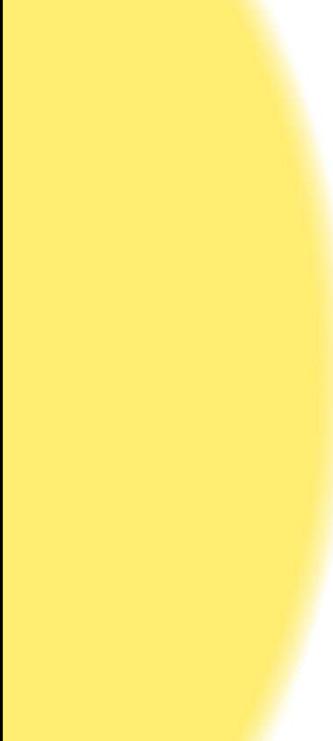


- Le DMP ou PGD est un document de quelques pages, qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée.
- Il permet de définir et de mettre en place les actions indispensables pour gérer les données qui vont être collectées ou produites au cours d'un projet. Il explique comment sont gérées les données depuis leur création ou collecte jusqu'à leur partage et leur archivage.
- Le DMP est évolutif : sa rédaction démarre dès le début d'un projet, mais il va être mis à jour régulièrement en fonction de l'évolution du projet. Vous n'avez pas à répondre à toutes les questions du DMP dans sa phase initiale mais il est bon de réfléchir à tous les points abordés dans le DMP. Ce document est là pour faciliter la gestion des données de recherche de votre projet.
- Il permet de **se poser les bonnes questions** pour rendre les données **FAIR**.

### Ressource :

- DoRANum. Cours introductif sur le Plan de Gestion de Données (PGD).  
<https://doi.org/10.13143/T3J4-VN03>

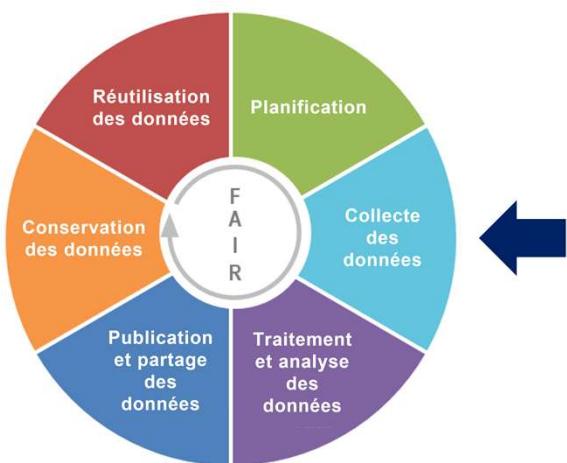
**Source des images :** Cocomaterial [CC0 1.0 (Public Domain Dedication)]



## **Étape 2 : Création, collecte et description des données de recherche**

20

## Place dans le cycle de vie des données



Il s'agit de la deuxième étape du cycle de vie des données.

**Source de l'image :** d'après *Research data lifecycle – UK Data Service*.

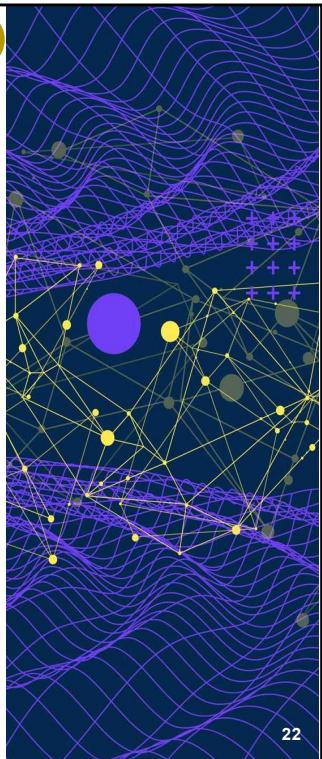
## Description des données de recherche



Produites ou recueillies



Préexistantes



Selon le projet, les données de la recherche peuvent être :

- **Produites ou recueillies** : ce sont des données créées spécifiquement durant le projet. Elles sont élaborées, générées lors d'activités de recherche (collectes sur le terrain, observations, des mesures...).
- **Préexistantes** : ce sont des données déjà existantes (provenant de corpus, archives...) qui sont utilisées pour le projet. Les données utilisées peuvent avoir été recueillies initialement dans un autre contexte que celui de la recherche mais elles sont utilisées comme données de recherche dans le cadre du projet.

Selon que les données sont produites ou préexistantes, leur description ne contiendra pas les mêmes informations.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

# Préparation et documentation des données

Dans le PGD, il faudra indiquer de manière précise **quelles méthodes** sont utilisées pour recueillir ou produire les données :

**Dans le cas de données préexistantes :**

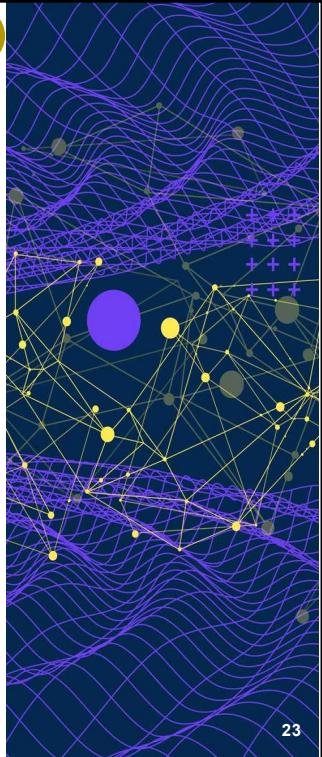
- leur provenance (corpus, archives...)
- sur quels critères elles ont été sélectionnées
- les conditions préexistantes de réutilisation de ces données

**Dans le cas de données produites ou recueillies :** (observations, mesures, etc.) :

- le contexte de création
- les méthodes utilisées
- les protocoles suivis ou établis
- les contrôles qualité mis en place



Attention aux **données sensibles, personnelles ou confidentielles** : prendre les précautions nécessaires afin de respecter les règles juridiques et éthiques en vigueur



Il est important de **bien documenter les données** de manière à ce qu'elles soient faciles à trouver et réutilisables.

Ces informations devront apparaître dans les métadonnées et/ou dans un fichier Readme. Elles sont indispensables pour que d'autres chercheurs puissent reproduire les résultats (reproductibilité et réplicabilité).

Dans le cas de **données préexistantes**, il faudra par exemple indiquer :

- Leur provenance
- Les critères selon lesquels elles ont été sélectionnées
- Les conditions préexistantes de réutilisation de ces données

Dans le cas de **données produites ou recueillies**, il faudra indiquer :

- le contexte de création
- les méthodes utilisées
- les protocoles suivis ou établis
- les contrôles qualité mis en place

Attention, en revanche en ce qui concerne les données personnelles (en particulier les données sensibles), et les données confidentielles : des précautions particulières sont à prendre.

Définitions à garder (si on nous pose la question) :

*La **reproductibilité** est la capacité, par une équipe différente, de reproduire une expérience, sans se fier au dispositif expérimental et aux codes logiciels développés par l'équipe d'origine.*

*La **réPLICABILITÉ** est la capacité, par une équipe différente, de reproduire une expérience en ré-utilisant le même dispositif expérimental décrit (y compris les codes logiciels).*

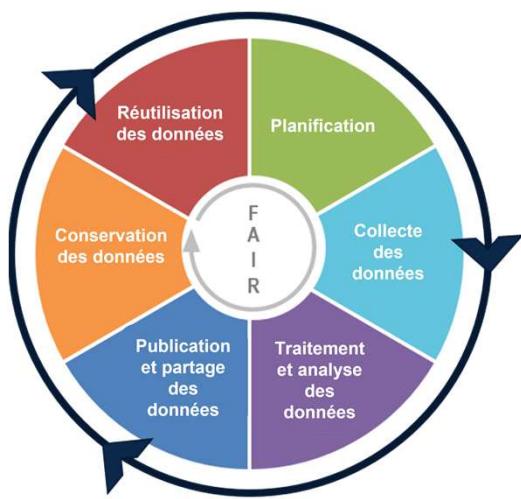
**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)



## Aspects juridiques et éthiques

24

## Place dans le cycle de vie des données



Les aspects juridiques et éthiques sont à considérer tout au long du cycle de vie des données.

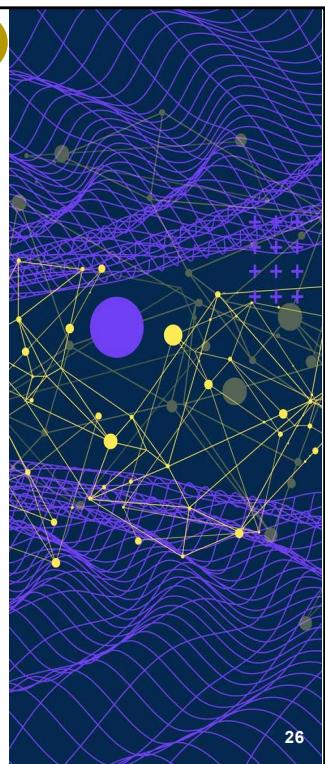
## Droits et obligations du chercheur

Cette étape est cruciale car elle détermine la latitude dont le chercheur disposera ensuite pour publier, diffuser et communiquer ses données et les résultats de ses recherches

Dès le début du projet, au moment de la collecte et de la production des données, le chercheur doit être vigilant concernant ses droits et obligations

**Exemples:**

- Dans le cas d'une interview ou de prises de son ou de vue, il doit recueillir le consentement écrit des personnes concernées
- Dans le cas de consultation de données d'archives, quels sont les droits afférents ?
- Dans le cas de collecte d'objets archéologiques, quels sont les droits liés au pays de collecte ?



### Ressource :

*Delplanque Catherine, Lamrini Nawale, Leclère Fabrice, Maurel Lionel, et al.*

*Guide : Règlement Général pour la Protection des Données. 2019.*

<https://direction-recherche.parisnanterre.fr/accueil/rgrpdc-fiches-pratiques-rgrpdc-recherche>

## Propriété des données

Accompagnement par un juriste recommandé

Règle générale



Si partenariat

- Données considérées comme propriété de l'établissement de tutelle des producteurs de données
- Nécessité d'établir au préalable un **accord de consortium**
- Cas de collaborations entre secteurs public et privé (déploiement industriel, commercialisation...)
- Cas de collaborations internationales

Penser à préciser la propriété des données et les responsabilités dans le PGD



**Attention, le PGD n'a aucune valeur juridique.**

« Le régime juridique des données est particulier car il n'y a pas à proprement parler de droits de propriété intellectuelle sur les données. En effet ces droits portent uniquement sur des œuvres de l'esprit présentant un degré minimal d'originalité telles qu'une peinture, un livre ou un article de recherche. Le fait que la collecte des données ait demandé un investissement (humain, matériel, financier) n'est pas créateur de droit d'auteur sur les données elles-mêmes. »

Le terme « données de recherche » est un terme générique qui englobe une multitude de réalités différentes.

Derrière le terme « données », on peut retrouver une image, une interview, une base de données, un logiciel, des données « brutes », des données « traitées »...

Cette grande diversité des données fait qu'elles ne relèvent pas toutes des mêmes droits et que de nombreux types de données produites n'entrent pas dans le champ de la propriété intellectuelle.

Par exemple, des données « brutes » issues de mesures et de comptages ne sont pas considérées comme des œuvres de l'esprit au sens du code de la propriété intellectuelle. Elles ne relèvent pas du droit d'auteur et appartiennent à l'établissement qui a réalisé l'investissement humain et financier pour les produire.

En revanche, les données qui vont donner lieu à un **dépôt de brevet ou à une déclaration d'invention** relèvent du champ de la propriété industrielle et seront protégées à ce titre.

De même, des jeux de données comprenant des photographies originales, c'est-à-dire portant l'empreinte de l'originalité de leur auteur, ou des extraits d'œuvres, vont relever de la **propriété intellectuelle** et du droit d'auteur.

Dans la cadre d'un partenariat (entre établissements, entre partenaires internationaux, entre secteurs public et privé), il sera nécessaire de définir au préalable qui détient les droits de propriété intellectuelle (par exemple via un accord de consortium ou autres types de contrats).

**Sources :**

- *Propriété des données | LabRADO.R.* (s. d.). Consulté 13 janvier 2025 : <https://labrador.sorbonne-universite.fr/cadre-juridique/propriete-des-donnees>
- *Comprendre la propriété des données dans un contexte d'ouverture | SOcle : Clés pour la Science Ouverte.* (s. d.). Consulté 13 janvier 2025 : <https://socle.univ-rennes2.fr/vos-besoins/comprendre-propriete-donnees-dans-contexte-douverture>
- *Diffuser et réutiliser en respectant la propriété intellectuelle et le droit d'auteur | SOcle : Clés pour la Science Ouverte.* (s. d.). Consulté 13 janvier 2025 : <https://socle.univ-rennes2.fr/vos-besoins/diffuser-reutiliser-respectant-propriete-intellectuelle-droit-dauteur>
- *MOOC science ouverte alliance sorbonne université* : <https://www.fun-mooc.fr/fr/cours/la-science-ouverte/>

**Pour aller plus loin :**

- *La titularité des droits de propriété intellectuelle des personnels de recherche.* (s. d.). Ouest Valorisation. Consulté 13 janvier 2025, à l'adresse <https://www.ouest-valorisation.fr/dossiers-d-experts/la-titularite-des-droits-de-propriete-intellectuelle-des-personnels-de-recherche/>

# Obligations/Recommandations de diffusion

Principe  
« Aussi ouvert  
que possible,  
aussi fermé que  
nécessaire »

## Obligations :

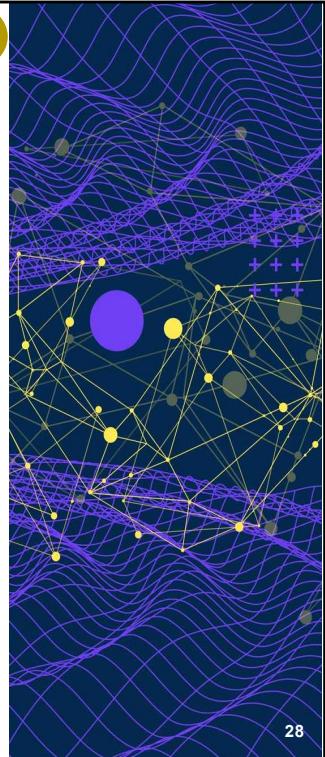
Suivant le financeur du projet, il peut être obligatoire de diffuser ses données et de rédiger un Plan de Gestion des Données (PGD)



## Recommandations :

Certaines institutions peuvent émettre des recommandations précises en matière de diffusion et de partage des données produites.

Exemples : Cirad, INRAe



## Ressources :

- Becard Nicolas, Castets-Renard Céline, Chassang Gauthier, Dantant Martin, Freyt-Caffin Laurence, Gandon Nathalie, Martin Caroline, Martelletti Andrea, Mendoza-Caminade Alexandra, Morcrette Nathalie, Neirac Claire. Ouverture des données de recherche. Guide d'analyse juridique en France. Décembre 2017.*  
<http://dx.doi.org/10.15454/1.481273124091092E12>
- Bordignon Frédérique, Boistel Romain, Du Pasquier Delphine. Qui a les droits, quelles obligations ?*  
[https://espacechercheurs.enpc.fr/sites/default/files/logigramme\\_a\\_plat\\_2.pdf](https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat_2.pdf)
- Ginouvès Véronique, Gras Isabelle, et al. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques.* <https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques>

## Communicabilité des données



La communicabilité des données peut être conditionnée par

- la nature ou le type des données
- l'origine des données
- leur (ré)utilisation

Elle peut être empêchée temporairement (embargo) ou définitivement

Toute restriction doit être expliquée dans le PGD

Communication obligatoire pour certaines disciplines

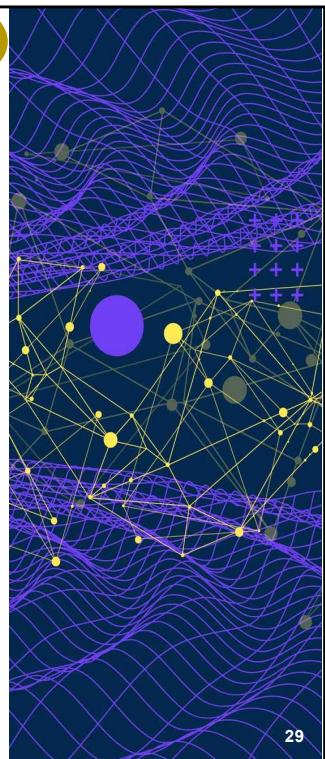
- Données géographiques
- Données environnementales...

Communication sous conditions

- Données protégées par le droit d'auteur ou par contrat
- Données personnelles
- Statistiques...

Communication interdite par principe

- Secrets professionnels
- Secrets défense
- Sécurité de l'établissement...



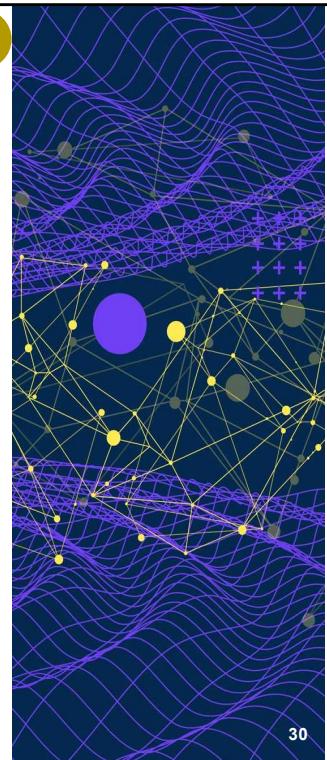
La communicabilité des données est réglementée. Elle peut être conditionnée par : la nature ou le type des données, l'origine des données, leur (ré)utilisation. Elle peut être empêchée temporairement (embargo) ou définitivement. Toute restriction doit être expliquée dans le PGD. Par exemple, les données environnementales doivent obligatoirement être diffusées (bien sûr, il existe toujours des exceptions). Les données personnelles peuvent être diffusées mais sous certaines conditions (par exemple, si elles sont anonymisées). Les données qui relèvent du secret défense sont interdites de communication.

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

# Éthique des données de recherche



- Dans le cas de données devant respecter des **règles d'éthique** particulières, se référer aux normes, chartes, déclarations, codes, politiques établis dans son domaine...
- Recourir à un **comité d'éthique**, si besoin



cnrs Inist

**L'éthique** nous invite à réfléchir aux valeurs qui motivent nos actes et à leurs conséquences et fait appel à notre sens moral et à celui de notre responsabilité.

**La déontologie** réunit les devoirs et obligations imposés à une profession, une fonction ou une responsabilité.

**L'intégrité scientifique** concerne la « bonne » conduite des pratiques de recherche. Elle désigne l'ensemble des règles et valeurs qui doivent régir l'activité de recherche pour en garantir le caractère honnête et scientifiquement rigoureux. Le respect de ces règles est une condition indispensable du maintien du lien de confiance accordée par la société aux acteurs de la recherche.

## Sources :

- CNRS, *Centre national de la recherche scientifique. Responsabilité de recherche*. 8 janvier 2024. <https://www.cnrs.fr/fr/le-cnrs/responsabilites/responsabilite-de-recherche>
- ANR. *L'intégrité scientifique*. <https://anr.fr/fr/lanr/engagements/lintegrite-scientifique/>

Des normes, chartes, déclarations, codes et politiques en éthique et en intégrité scientifique encadrent les pratiques pour l'ensemble de nombreux acteurs internationaux de la recherche.

**Ressources :**

- ANR, Agence nationale de la recherche. *Politique en matière d'éthique et d'intégrité scientifique*. Août 2014.  
<https://anr.fr/fileadmin/documents/2014/Politique-ethique-integrite-scientifique-aout-2014.pdf>
- Ginouvès Véronique, Gras Isabelle et al. *La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques* :  
<https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques>
- Questions éthique et droit en SHS : <https://ethiquedroit.hypotheses.org/>
- DoRANum. Qu'est-ce que l'intégrité scientifique ?  
<https://doi.org/10.13143/GSKB-PK20>

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## En résumé

Pour être sûr que les données sont accessibles et réutilisables,  
veillez à respecter

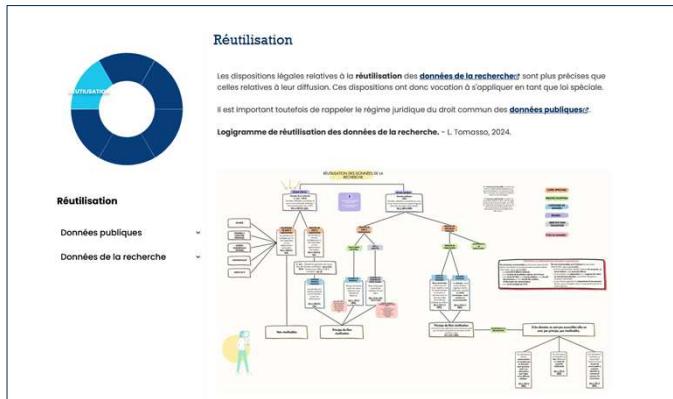
Les aspects juridiques	Les aspects éthiques
<ul style="list-style-type: none"><li>• Propriété intellectuelle</li><li>• Accords de consortium</li><li>• Obligations de diffusion</li><li>• Communicabilité des données</li><li>• ...</li></ul>	<ul style="list-style-type: none"><li>• Comité d'éthique</li><li>• Normes</li><li>• Chartes</li></ul>  

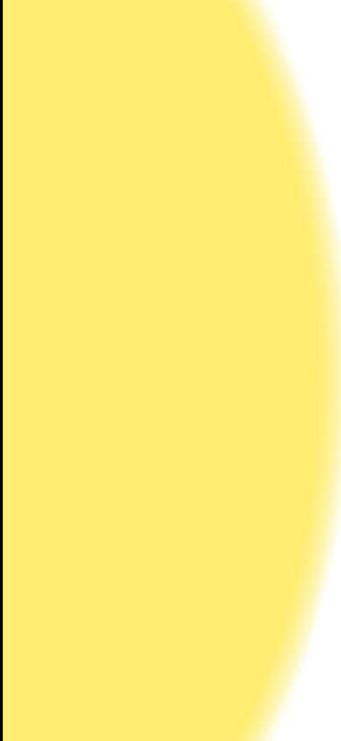
# Ressources juridiques



Centre de ressources  
ressources juridiques des données de la recherche  
■ ■ Recherche Data Gouv

Le site **JurisDoR** (centre de ressources de Recherche Data Gouv) diffuse les ressources juridiques nécessaires à la gestion des données de la recherche.

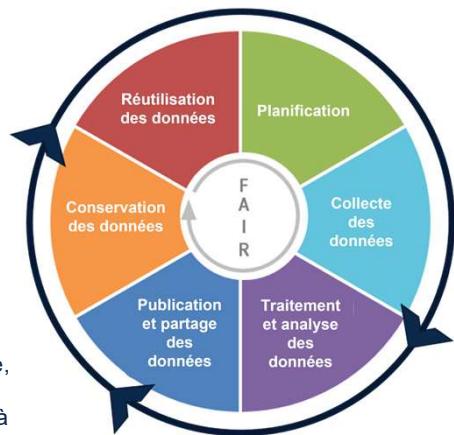




## Métadonnées

33

## Place dans le cycle de vie des données



Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet

Au moment du partage puis de l'archivage pérenne, des métadonnées spécifiques seront à renseigner

Schéma : DoRANum  
D'après UK data service – Research data lifecycle



34

Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet.

## Définition – utilisation des métadonnées

- Les métadonnées permettent de décrire plus précisément les données
- Elles sont à renseigner au fur et à mesure pour chaque jeu de données
- La boîte de conserve = jeu de données / l'étiquette = métadonnées



Sans métadonnées



Avec métadonnées



35

Les métadonnées sont **toutes les informations annexes qui décrivent les données pour faciliter leur compréhension et leur réutilisation.**

Par exemple, les métadonnées d'une boîte de conserve sont toutes les informations que l'on trouvera sur l'étiquette de la boîte : nom du produit, marque, liste des ingrédients, nombre de calories, teneur en glucides, en sel...

En contexte de recherche scientifique, il est important de renseigner les métadonnées au fur et à mesure pour chaque jeu de données, pour toutes les étapes du cycle de vie des données.

### Ressource :

*DoRANum. Cours introductif sur les métadonnées.*

<https://doi.org/10.13143/T3J4-VN03>

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Métadonnées embarquées et enrichies

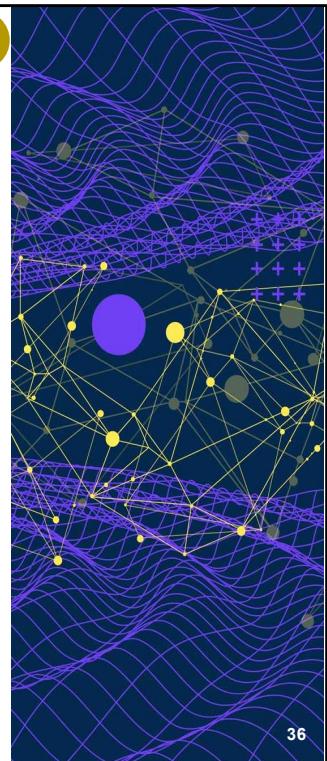
Compléter les métadonnées embarquées par des métadonnées enrichies

Métadonnée embarquée :  
produite automatiquement  
par les appareils ou logiciels

Métadonnée enrichie :  
ajoutée par l'auteur

Ex : date, format, données GPS,  
type d'appareil, calibrage  
technique, etc.

Ex : mots-clés, sujet, auteur,  
laboratoire ou organisme, nom du  
projet, licence, etc.



On distingue deux types de métadonnées :

- Les **métadonnées embarquées** sont les métadonnées produites automatiquement par des appareils (de mesure, de prise de vue ou de son...) ou des logiciels
- Les **métadonnées enrichies** sont les métadonnées ajoutées manuellement par l'auteur (ex : mots-clés, licence, organisme d'affiliation, protocole utilisé...).

## Rôle des standards de métadonnées

Selon le type de données, les informations qui aident à décrire les données peuvent être spécifiques.

On ne décrira pas de la même manière un vêtement, une maison ou un fruit :



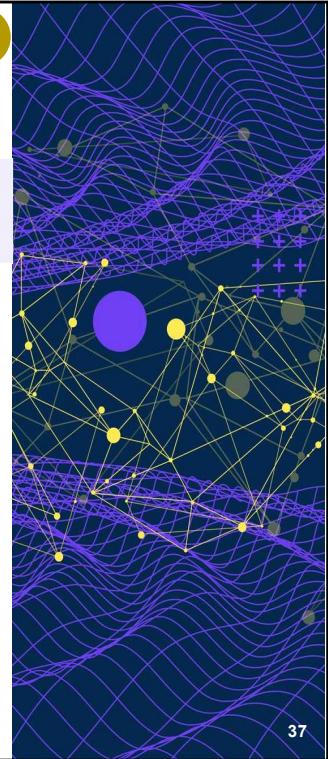
- Matière
- Température de lavage
- Pays de fabrication
- Marque
- ...



- Surface (en m<sup>2</sup>)
- Nombre de pièces
- Classe énergétique
- Garage ou pas
- ...



- Variété
- Pays de provenance
- Bio ou pas
- ...



Selon le type de données, les informations qui aident à décrire les données peuvent être spécifiques.

En effet, on ne décrira pas de la même manière un vêtement, une maison ou un fruit.

Pour un vêtement, on parlera de : sa matière, sa température de lavage, son pays de fabrication, sa marque...

Pour une maison, on évoquera : la surface (en m<sup>2</sup>), le nombre de pièces, la classe énergétique, la présence ou pas d'un garage...

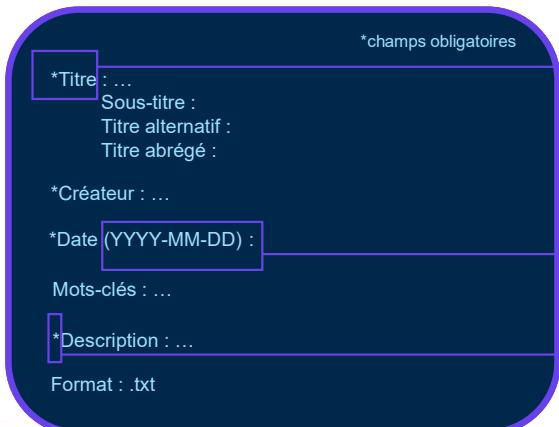
Pour un fruit, on pourra parler de la variété, du pays de provenance, de sa culture bio ou pas.

De la même manière qu'on décrira différemment un vêtement, une maison ou un fruit... on décrira différemment la photographie d'une peinture pariétale et les traits phénotypiques d'un arbre.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Standard de métadonnées ?

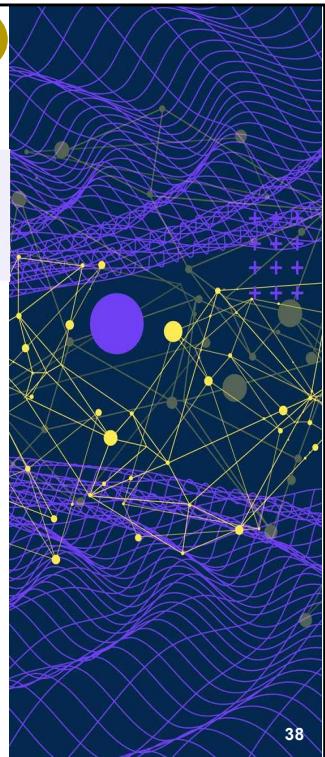
Un standard est un schéma qui a été adopté comme modèle par un ensemble d'utilisateurs : il est reconnu, normalisé et utilisé à grande échelle



Quoi

Comment : par exemple,  
norme ISO 8601 pour la date

Champ obligatoire



CNRS Inist

Mais alors comment décrire au mieux ses données en s'adaptant à leurs spécificités ? La pratique recommandée est de s'appuyer sur un standard de métadonnées pour décrire les données.

Pour schématiser simplement ce qu'est un standard de métadonnées, disons que c'est un plan ou formulaire (adopté par une communauté) qui indique :

- quelles métadonnées, c'est-à-dire quelles informations renseigner,
- dans quel ordre,
- comment remplir les champs de métadonnées (par exemple : date de création en utilisant la norme ISO 8601)
- quels champs sont obligatoires.

# Standards disciplinaires



**Data Documentation Initiative**  
Enquêtes



**Darwin Core**  
Données de biodiversité



2.3. méthode

2.3.1 collecte des données  
2.3.1.1

2.3.1.6. mode de collecte

Questionnaire à distance,  
interviews en face à face,  
entretien téléphonique...

Règne :  
Embranchement :  
Classe :  
Ordre :  
Famille :  
Genre :  
...

Choix parmi les 7 règnes :  
animal, végétal, bactéries...

Une description qui suit les rangs  
taxonomiques pour la classification  
hiérarchique du monde vivant

CNRS Inist

39

Dans la définition d'un standard de métadonnées, on retrouve l'idée qu'il s'agit d'un modèle **adopté par une communauté**. La plupart du temps il s'agit de **communautés disciplinaires**. Ceci permet d'avoir des métadonnées qui sont **adaptées à la discipline/aux types de données de la discipline**.

Par exemple :

- Data Documentation Initiative (DDI) est un standard adapté aux enquêtes en sciences sociales, économiques et politiques. Une partie de ses champs concerne ainsi la méthode de collecte des données. Par exemple, ce standard prévoit un champ de métadonnées dédié au mode de collecte des données. Il faut préciser dans ce champ si les données ont été collectées par : questionnaires envoyés par mail, interviews en face à face, entretien téléphonique...
- Par comparaison, Darwin Core permet de décrire des données en biodiversité. Ce standard prévoit donc de renseigner plusieurs rangs hiérarchiques de taxonomie : règne, famille, genre, espèce...

Gardez à l'esprit que, renseigner les métadonnées, c'est renseigner ce qu'il importe de savoir pour comprendre les données.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Standard et schéma de métadonnées

Utiliser un standard de sa discipline ou adapté à ses besoins et, s'il n'en existe pas, créer un schéma de métadonnées

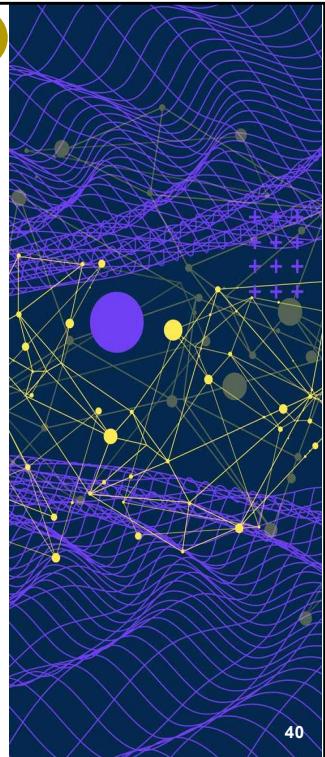


### Standard

Un standard est un schéma qui a été adopté comme modèle par un ensemble d'utilisateurs : il est reconnu, normalisé et utilisé à grande échelle  
⇒ Commun et interopérable

### Schéma

C'est l'organisation des métadonnées selon un plan créé pour un besoin spécifique  
⇒ Unique et personnalisé



Si aucun standard de métadonnées existant ne répond à vos besoins de description, il est tout à fait envisageable **de créer votre propre schéma de métadonnées**.

L'avantage étant que, conçu pour les besoins spécifiques d'un projet, il est très adapté aux données pour lesquelles il a été créé. L'inconvénient est, qu'étant spécifique, il perd en interopérabilité.

En effet, opter pour un schéma de métadonnées plutôt qu'un standard rend vos données beaucoup moins interopérables et moins FAIR. Seules les métadonnées renseignées selon un standard seront interopérables. Les métadonnées spécifiques le seront moins voire pas du tout. Pour ces questions, **n'hésitez pas à vous faire aider par un documentaliste ou un informaticien**.

À garder à l'esprit : l'utilisation d'un standard de métadonnées, notamment disciplinaire, est un élément clé pour atteindre un haut degré de respect des principes FAIR :

- **Facile à trouver** : une donnée n'est souvent trouvable que par les éléments de métadonnées indexés dans le moteur de recherche consulté.
- **Accessible** : notamment grâce aux métadonnées.
- **Interopérable** : grâce au standard commun qui facilite les traitements informatiques.
- **Réutilisable** : grâce à la provenance décrite dans les métadonnées, la licence attribuée et au recours à un standard disciplinaire.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

**Ressource :**

- DoRANum. *Les schémas de métadonnées.* <https://doi.org/10.13143/J8BC-0Z74>

## Exemples de standards de métadonnées

### Dublin Core

- Standard interdisciplinaire : description des ressources numériques

### DataCite Metadata Schema

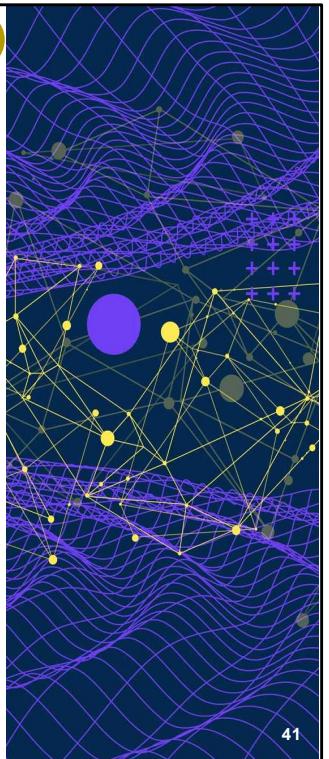
- Standard lié à l'attribution d'identifiants pérennes DOI

### DDI (Data Documentation Initiative)

- Sciences sociales, comportementales et économiques

### CSMD-CCLRC Core Scientific Metadata Model

- Sciences structurelles (chimie, science des matériaux, sciences de la terre, biochimie)



Exemples de standards de Métadonnées :

- **Dublin Core (interdisciplinaire)** : description des ressources numériques.  
<http://dublincore.org/>
- **DataCite Metadata Schema** : métadonnées enregistrées dans le DataCite Metadata Store lors de la création d'un DOI pour un jeu de données. Permet l'identification précise et cohérente des données à des fins de citation et de réutilisation. <https://schema.datacite.org/>
- **DDI (Data Documentation Initiative)** : sciences sociales, comportementales et économiques. <http://www.ddialliance.org/>
- **CSMD-CCLRC Core Scientific Metadata Model** : sciences structurelles (chimie, science des matériaux, sciences de la terre, biochimie)  
<http://icatproject-contrib.github.io/CSMD/>

## Exemples de standards de métadonnées

### DwC (Darwin Core)

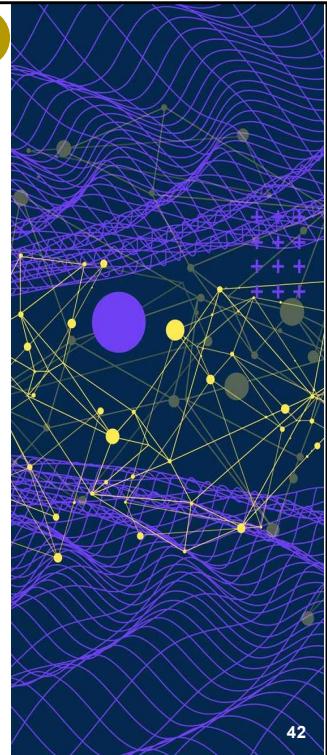
- Biodiversité

### EML (Ecological Metadata Language)

- Écologie

### MIDAS-Heritage

- Architecture



### Exemples de standards de Métadonnées :

- **DwC (Darwin Core)** : biodiversité. <http://rs.tdwg.org/dwc/>
- **EML (Ecological Metadata Language)** : très développé en écologie. En grande partie conçu pour décrire des ressources numériques. Il peut également être utilisé pour décrire des ressources non numériques telles que des cartes papier ou d'autres médias.  
<https://knb.ecoinformatics.org/external//emlparser/docs/index.html>
- **MIDAS-Heritage** : architecture. <https://historicengland.org.uk/images-books/publications/midas-heritage/>

### Ressources pour trouver des standards de métadonnées :

- *DoRANum. Les standards de métadonnées : pourquoi, lequel ?*  
<https://doi.org/10.13143/Y5PY-W521>
- *FAIRsharing.org. Répertoire de standards de métadonnées en Sciences de la Vie.*  
<https://fairsharing.org/search?fairsharingRegistry=Standard>
- *RDA, Research Data Alliance. Metadata Standards Catalog.*  
<https://rdamsc.bath.ac.uk/>
- *DCC, Digital Curation Centre. Disciplinary Metadata.*  
<https://www.dcc.ac.uk/guidance/standards/metadata>

Consulter également les informations fournies par les entrepôts de données sur les standards de métadonnées.

## Exemple multidisciplinaire : le Dublin Core

Élément simple	Définitions	Exemple d'élément spécifique
Title	nom de la ressource	
Subject	thème du contenu de la ressource	
Description	résumé, table des matières...	
Creator	auteur principal de la ressource	
Publisher	entité responsable de la diffusion de la ressource	
Contributor	co-auteurs associés à l'élaboration de la ressource	
Date	date de création ou mise à disposition	<u>Date Created</u> , <u>Date copyrighted</u> , <u>Date Valid</u> , <u>Date Available</u> , <u>Date Modified</u> , <u>Date Accepted</u> , <u>Date Submitted</u> , <u>Date Issued</u>
Type	nature du contenu : image, son texte...	
Format	format ou taille de la ressource	
Identifier	référence unique, DOI, URL, ISSN...	
Source	référence à une ressource à partir de laquelle la ressource actuelle a été dérivée ou créée	<u>Has Format</u> (les relations de transformation de format sont celles où une ressource a été dérivée d'une autre à l'aide d'une technologie de reproduction ou de reformatage qui n'est pas fondamentalement une interprétation mais une représentation.)
Language	langue originale de la ressource	
Relation	référence à une ressource apparentée	
Coverage	périmètre spatial et temporel	<u>isVersionOf</u> (les relations de version sont celles où une ressource est un état ou une parution historique d'une autre ressource par le même créateur)
Rights	informations sur les droits associés à la ressource	

Le Dublin Core est un **standard international et multidisciplinaire**.

Il **comporte 15 éléments** (= minimum exigé) ayant trait au contenu, à la propriété intellectuelle, à la version.

Ces 15 éléments restent cependant insuffisants pour représenter finement certaines données dans un univers dédié pour des utilisateurs et applications spécifiques. Une quarantaine d'éléments plus spécifiques, qui constituent le Dublin Core étendu, viennent ainsi préciser les éléments simples.

### Sources :

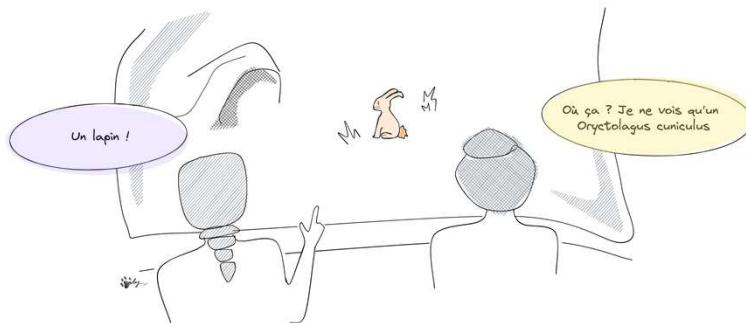
- DoRANum. Métadonnées, standards, formats : fiche synthétique.  
<https://doi.org/10.13143/VBJS-6288>
- Wikipédia. Dublin Core. [https://fr.wikipedia.org/wiki/Dublin\\_Core](https://fr.wikipedia.org/wiki/Dublin_Core)

# Enrichissement des métadonnées

- Utiliser des vocabulaires contrôlés disciplinaires utilisés couramment (lexiques, thesaurus, ontologies...)

- Exemples :

- Codex de médicaments
- Classifications taxonomiques
- Nomenclature internationale des formules chimiques



- Cela augmentera la capacité des données à être combinées avec d'autres données (interopérabilité)

Pour être FAIR, il faut s'appuyer sur des vocabulaires contrôlés (lexiques, thesaurus, ontologies...), de préférence disciplinaires, pour sélectionner les termes à indiquer dans les champs de métadonnées.

« Le **thesaurus** est un ensemble organisé de termes contrôlés et normalisés représentant les concepts d'un domaine de connaissance. Les termes sont reliés entre eux par des relations de synonymie (terme équivalent), de hiérarchie (terme générique et terme spécifique) et d'association (terme associé) ; chaque terme appartient à une catégorie ou domaine. » (*Wikipedia. Thésaurus documentaire*.)

[https://fr.wikipedia.org/wiki/Th%C3%A9saurus\\_documentaire](https://fr.wikipedia.org/wiki/Th%C3%A9saurus_documentaire))

« L'**ontologie** est un ensemble organisé de termes/concepts avec des relations sémantiques variées décrivant un domaine de connaissance. Exemple : FOAF (FriendOfAFriend). (*Ecole thématique E-Envir 2021. Interoperable & Reuse. Introduction aux concepts clefs et immersion. 2 au 5 novembre 2021.* [https://e-envir-21.sciencesconf.org/data/pages/6\\_E\\_ENVIR21\\_INTEROP\\_REUSE.pdf](https://e-envir-21.sciencesconf.org/data/pages/6_E_ENVIR21_INTEROP_REUSE.pdf))

« La **taxonomie** ou taxinomie est une branche des sciences naturelles qui a pour objet l'étude de la diversité du monde vivant. Cette activité consiste à décrire et circonscrire en termes d'espèces les organismes vivants et à les organiser en catégories hiérarchisées appelées taxons. (*Wikipedia. Taxonomie.* <https://fr.wikipedia.org/wiki/Taxonomie>)

« Par extension, [la taxonomie est une] classification, suite d'éléments formant des listes qui concernent un domaine, une science. » (*Dictionnaire Larousse*.)

<https://www.larousse.fr/dictionnaires/francais/taxinomie/76893>

## Utilité des métadonnées

Les métadonnées sont utiles pour :

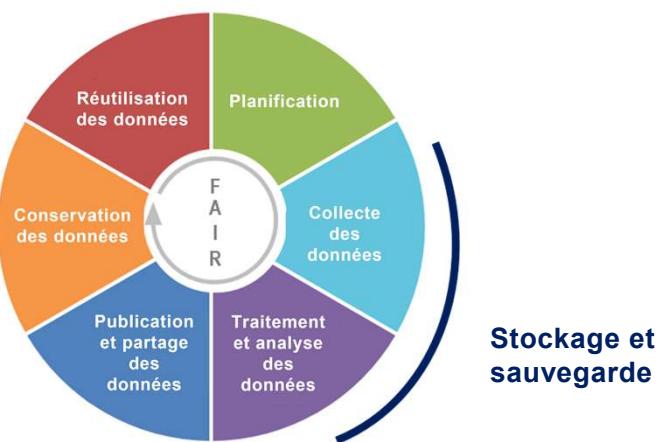
- Comprendre l'origine des données et leur contexte de création ou de collecte
- Améliorer le moissonnage par les machines (moteurs de recherche)
- Garantir l'interopérabilité
- Connaitre les conditions de réutilisation et de partage des données
- Accéder à des informations très utiles lorsqu'on ne peut pas partager ses données ou lors de la suppression des données obsolètes

Quoiqu'il arrive aux données, les métadonnées resteront toujours accessibles. Il est important de mentionner que les données ont été supprimées et pourquoi.



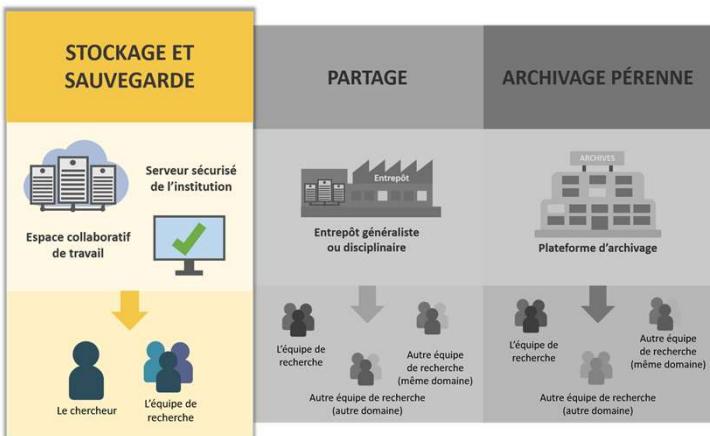
## **Étape 3 : Traitement et analyse Stockage sécurisé durant le projet**

## Place dans le cycle de vie des données



Les questions de stockage et de sauvegarde des données interviennent aux étapes de collecte et de traitement des données.

## Stockage et sauvegarde des données



Le **stockage** et la **sauvegarde** sécurisés des données se font durant le projet.

L'objectif est de garantir la sécurité des données et d'en faciliter l'accès pour l'ensemble des collaborateurs du projet.

### Ressource :

- *DoRANum. Stockage, partage et archivage : quelles différences ?*  
<https://doi.org/10.13143/5DAX-QP58>
- *DoRANum. Stocker ses données de façon sécurisée.*  
<https://doi.org/10.13143/Z0GE-NC29>

## Mesures de sauvegarde (stockage)

- Sélectionner les données
- Appliquer la **règle du 3-2-1**



Garder **3** exemplaires des données  
sur **2** supports ou technologies différents,  
dont **1** se trouve hors site

- Déterminer l'hébergement
- Estimer la volumétrie des données et la durée de conservation

Organiser et planifier la sauvegarde des données :

- À chaque point d'étape du projet, sélectionner les données à sauvegarder, à supprimer  
Dans l'idéal, dupliquer et stocker les données à différents endroits sur différents supports  
en veillant à bien gérer les versions :
- Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement
- Permet de revenir à une version antérieure si besoin.

Le mieux est de suivre la règle du 3-2-1, c'est-à-dire : garder 3 exemplaires des données, sur 2 supports ou technologies différents dont 1 se trouve hors site.

Déterminer l'hébergement : serveurs locaux (machines virtuelles), cloud institutionnel avec accès sécurisé...

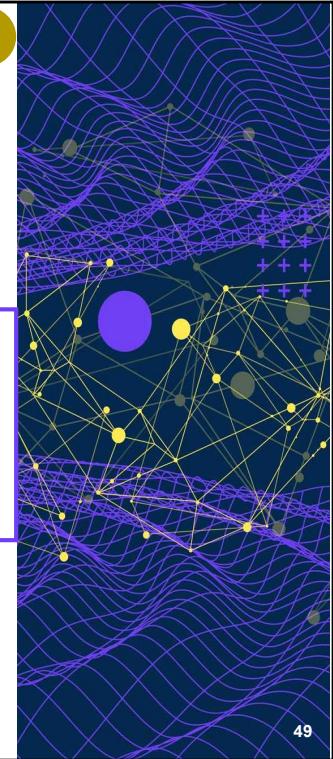
Estimer la volumétrie des données et la durée de conservation

**Contactez les services de votre institution** pour l'hébergement, la volumétrie, la durée de conservation.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

**Sources :**

- Université de Lille Sciences et Technologies. *Problématique des sauvegardes – Journées Mathrice (sept 2017)*
- DoRANum. *La sauvegarde 3-2-1*. <https://doi.org/10.13143/1GDH-TK61>



## Exemples d'espaces de stockage



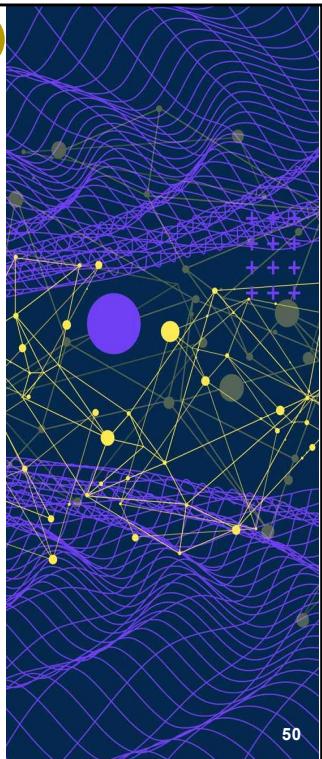
À privilier

- Plateformes institutionnelles
- Infrastructures disciplinaires



À limiter

- Clé USB
- Disque dur
- Solutions commerciales



Exemples **Plateformes institutionnelles ou infrastructures disciplinaires** :

- Huma-Num Box : <https://documentation.huma-num.fr/humanum-box/>
- Stockage données recherche (PETA) à l'Université de Lorraine :  
<https://numerique.univ-lorraine.fr/catalogue-des-services/stockage-donnees-recherche-peta>
- Espaces de stockage du CC-IN2P3 : <https://doc.cc.in2p3.fr/fr/Data-storage/storage-areas.html>

Exemples de **solutions commerciales** : Google Drive, Dropbox...

**Source des images** : Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Formats de fichiers



### Formats ouverts et non propriétaires

Opter pour des formats de fichiers les plus ouverts possible (non propriétaires), standardisés et pérennes

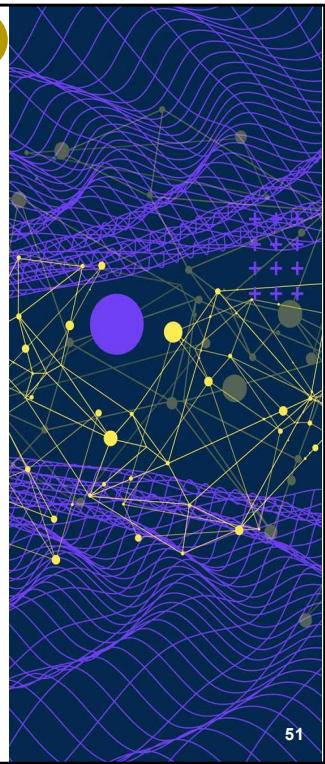
- Exemples :
- Privilégier .csv à .xlsx
  - Privilégier .odt à .docx
  - Privilégier .jpg à .psd
  - Privilégier .zip à .rar

### Choix du format



Le choix d'un format peut être guidé par :

- les recommandations de son institution
- les usages de la communauté scientifique de la discipline
- les logiciels ou équipements utilisés



### Ressource :

DoRANum. Format ouvert ou fermé ? <https://doi.org/10.13143/MCWQ-QS64>

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Nommage des fichiers

La fiabilité d'accès passe par un nommage unique et précis des fichiers de données :



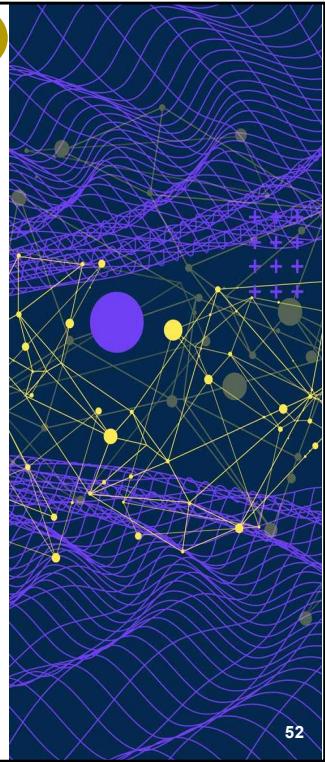
Bonnes pratiques

- 30 caractères maximum
- Nom du projet
- Numéros de versions le cas échéant
- Dates au format ISO : AAAA-MM-JJ
- ...



A éviter

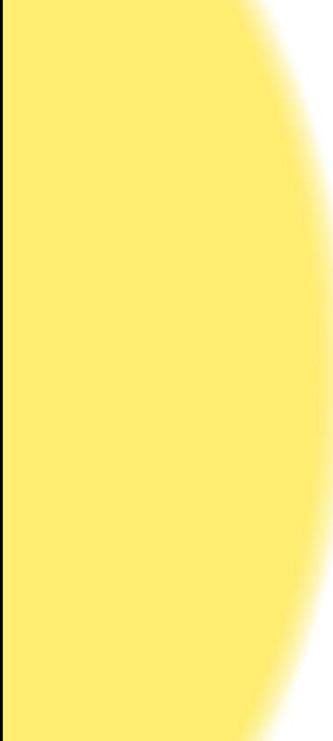
- Pas de caractères spéciaux ou accentués du type ùéàç+'@°[ ] :</\* »& !\$ ...
- Séparateurs : pas d'espace, pas de mots vides, éventuellement Majuscules ou underscore \_
- Pas de dénomination vague : divers, autres, à classer...



### Source :

Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine. *Guide de bonnes pratiques : Gestion et valorisation des données de recherche*. 23 février 2016.  
<https://hal.science/hal-01275841/>

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)



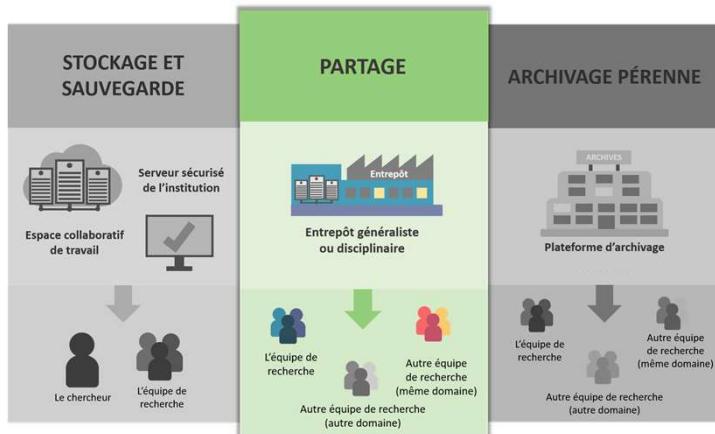
## **Étape 4 : Partage, diffusion des données Dépôt dans un entrepôt**

## Place dans le cycle de vie des données



La question du partage des données intervient normalement vers la fin du cycle de vie des données.

## Partage, diffusion des données



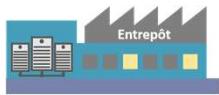
Le **partage** consiste à déposer les données dans un entrepôt de données afin de les rendre accessibles facilement et de permettre leur réutilisation par des chercheurs du même domaine ou d'un autre domaine, selon les principes FAIR et sur le court et le moyen terme. Le partage se fait souvent à l'issue du projet.

### Ressource :

- *DoRANum. Stockage, partage et archivage : quelles différences ?*

<https://doi.org/10.13143/5DAX-QP58>

## Dépôt



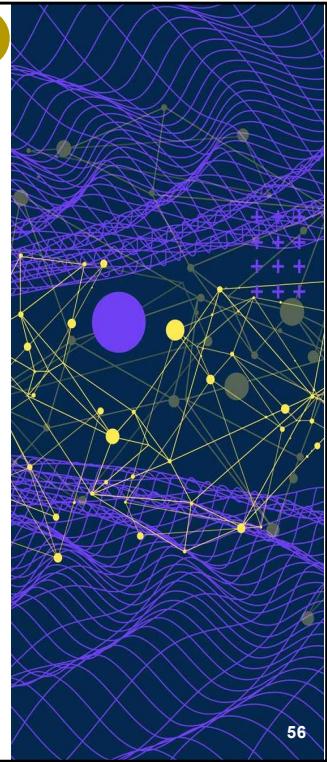
Software Heritage

Déposer les **données**  
(et métadonnées  
associées)

Déposer les  
**codes sources** liés  
dans l'archive  
Software Heritage

5 à 10 ans

Plus de 10 ans



Il est recommandé de déposer ses données dans un entrepôts et de renseigner au maximum les métadonnées.

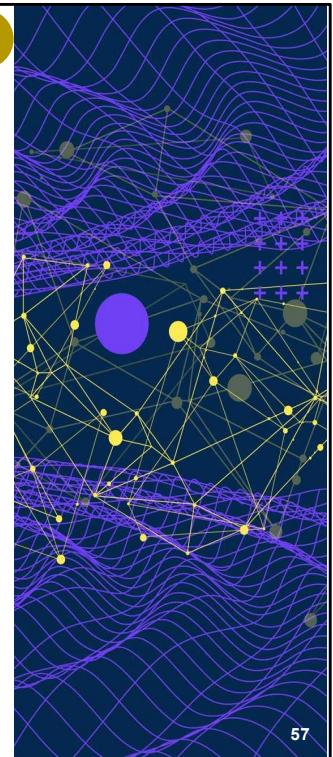
Pour les logiciels, il est possible de déposer les codes sources dans Software Heritage, archive de logiciels (<https://www.softwareheritage.org/?lang=fr>).

## Préparation des données pour le partage



### Check-list

- Sélectionner les données à partager
- Estimer la volumétrie des données
- Vérifier la compatibilité et l'interopérabilité des formats de données
- Migrer si possible vers un format ouvert
- Préparer si nécessaire les codes sources
- Compléter et enrichir les métadonnées
  - Si ce n'est pas déjà fait, choisir un standard de métadonnées
  - S'il n'en existe pas d'adapté, créer un schéma de métadonnées
  - Compléter les champs pour chaque jeu de données



Avant de déposer les données dans un entrepôt, il faut les préparer :

- Sélectionner les données à partager
- Estimer la volumétrie des données
- Vérifier la compatibilité et l'interopérabilité des formats de données (I/R)
- Migrer si possible vers un format ouvert (I/R)
- Préparer si nécessaire les codes sources qui permettront de lire et traiter les données (A/I/R)
- Compléter et enrichir les métadonnées (en fonction de l'entrepôt choisi) (F/A/I/R) :
  - Si ce n'est pas déjà fait, choisir un standard de métadonnées
  - S'il n'en existe pas d'adapté, créer un schéma de métadonnées
  - Compléter les champs pour chaque jeu de données, suivant le standard adopté.

Tous ces éléments contribuent aux principes FAIR.

### Ressource :

*DoRANum. Vérifier ses données de recherche.* [https://doranum.fr/depot-entrepos/verifier-donnees-recherche\\_10\\_13143\\_5rs6-4r06/](https://doranum.fr/depot-entrepos/verifier-donnees-recherche_10_13143_5rs6-4r06/)

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Choix de l'entrepôt de données

1

Privilégier un entrepôt **disciplinaire**

Exemples : [PANGAEA](https://www.pangaea.de/) (Environnement) ; [GenBank](https://www.ncbi.nlm.nih.gov/genbank/) (Génétique) ; [Ortolang](https://www.ortolang.fr/) (Linguistique)...

2

Si aucun entrepôt disciplinaire n'existe, choisissez un entrepôt **institutionnel**

Exemples : [CNRS Research Data](https://entrepot.recherche.data.gouv.fr/dataverse/cnrs) ; [Data INRAE](https://entrepot.recherche.data.gouv.fr/dataverse/inrae) ; [DOREL](https://entrepot.recherche.data.gouv.fr/dataverse/dorel-lorraine)...

3

En alternative, vous pouvez déposer vos données dans l'entrepôt national pluridisciplinaire [Recherche Data Gouv](https://entrepot.recherche.data.gouv.fr/)



Exemples d'entrepôts disciplinaires :

- **PANGAEA** (Environnement) : <https://www.pangaea.de/>
- **GenBank** (Génétique) : <https://www.ncbi.nlm.nih.gov/genbank/>
- **Ortolang** (Linguistique) : <https://www.ortolang.fr/>

Exemples d'entrepôts institutionnels :

- **CNRS Research Data** : <https://entrepot.recherche.data.gouv.fr/dataverse/cnrs>
- **Data INRAE** : <https://entrepot.recherche.data.gouv.fr/dataverse/inrae>
- **DOREL** : <https://entrepot.recherche.data.gouv.fr/dataverse/dorel-lorraine>

Entrepôt pluridisciplinaire :

**Recherche Data Gouv** : <https://entrepot.recherche.data.gouv.fr/>

## Exemples de critères de choix d'un entrepôt :

Entrepôt disciplinaire ou institutionnel

Type de données acceptées

Qualité des métadonnées

**Entrepôt de confiance - Certification**

Pérennité des métadonnées et des données

Génération d'un identifiant unique pérenne

Gestion des versions

Attribution d'une licence



Voici quelques critères pouvant aider à choisir un entrepôt.

Dans tous les cas, il est conseillé d'opter pour un entrepôt de **confiance**, voire **certifié**.

### Sources :

- RDA. *Entrepôts de données de confiance : critères de conformité.*  
<https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf>
- Science Europe. *Practical guide to the international alignment of research data management – Extended Edition.* 27 janvier 2021.  
<https://doi.org/10.5281/zenodo.4915862>

### Ressource :

DoRANum. *Les critères pour choisir un entrepôt de données.*

<https://doi.org/10.13143/ZQPB-9449>

## Annuaires et catalogues d'entrepôts

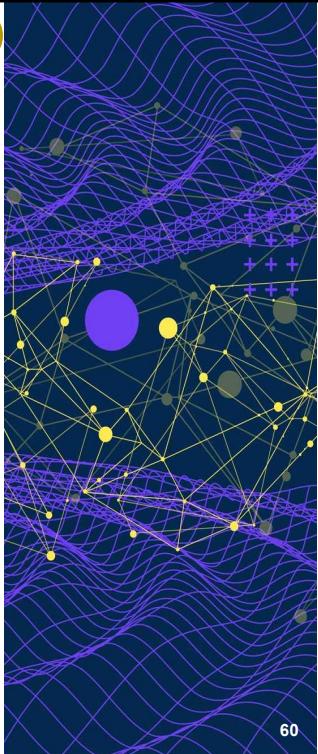
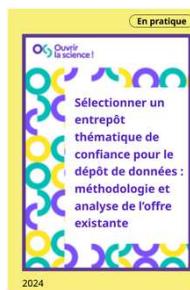
Les annuaires et catalogues permettent d'identifier un entrepôt disciplinaire qui conviendrait à vos besoins :



CNRS Inist



Le Comité pour la Science ouverte a établi une [liste d'entrepôts thématiques de confiance](#)



- **CatOPIDoR** : Catalogue pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche. Wiki des services dédiés aux données de recherche. Il recense des entrepôts français.  
[https://cat.opidor.fr/index.php/Entrep%C3%A8t\\_de\\_donn%C3%A9es](https://cat.opidor.fr/index.php/Entrep%C3%A8t_de_donn%C3%A9es)
- **Re3data** : annuaire recensant des entrepôts au niveau international et permettant de filtrer selon plusieurs critères (discipline, attribution d'un identifiant pérenne, types de données ou formats acceptés, préservation à long terme des données, choix de la licence, certification, etc.). <https://www.re3data.org/>
- **FAIRsharing** : ressource permettant d'identifier et citer des standards, des bases de données ou des entrepôts existant pour leurs données et leur discipline.  
<https://fairsharing.org/>
- **Collège Données de la recherche du Comité pour la Science Ouverte** : liste des entrepôts de confiance. Avril 2025. <https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-le-depot-de-donnees-methodologie-et-analyse-de-loffre-existante/>
- **OpenDOAR** : annuaire mondial d'entrepôts ou d'archives ouvertes en libre accès. La recherche et la navigation peuvent s'effectuer sur différents critères tels que le nom, la région du monde ou le pays et le logiciel. <https://v2.sherpa.ac.uk/opendoar/>
- **OAD** : Open Access Directory – Data repositories : liste d'entrepôts et de bases de données pour les données ouvertes. [https://oad.simmons.edu/oadwiki/Data\\_repositories](https://oad.simmons.edu/oadwiki/Data_repositories)

F
A
I
R

# Outil Cat OPIDoR

 Rechercher sur Cat OPIDoR Rechercher Demander un compte Se connecter

**Entrepôt de données**

Sommaire [masquer] Page Discussion Lire Voir le texte source Voir l'historique Outils ▾

Début Ajouter un nouveau service

- Sur quelles plateformes puis-je déposer et partager les données que j'ai produites au cours de mes recherches ?
- Existe-t-il un entrepôt français dans ma discipline de recherche ?
- Pour élargir votre recherche, vous pouvez consulter le registre re3data.

(+) Vous trouverez dans cette liste des entrepôts non portés par des structures françaises. Ils sont tous issus de la liste des entrepôts thématiques de confiance proposée par le Collège des Données de la recherche du Comité pour la science ouverte.

A ce jour, 187 entrepôts sont référencés dans Cat OPIDoR.

8-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [ ]

Vous pouvez utiliser les filtres afin d'affiner votre recherche :

Type d'entrepôt :  Disciplinaire  Institutionnel  Lié à un projet ou un équipement etc.  Multidisciplinaire  National

Domaine scientifique :  Sciences & Technologies  Sciences Humaines & Sociales  Vie & Santé

— Sciences Humaines & Sociales

— Sciences & Technologies

— Vie & Santé

Thématique/Mots clés : Sélectionner une valeur de filtre

Label ou certification :  OU  ET

Centre de référence thématique  CoreTrustSeal  CoRTecS  CoSO  ELIXIR Data Resources

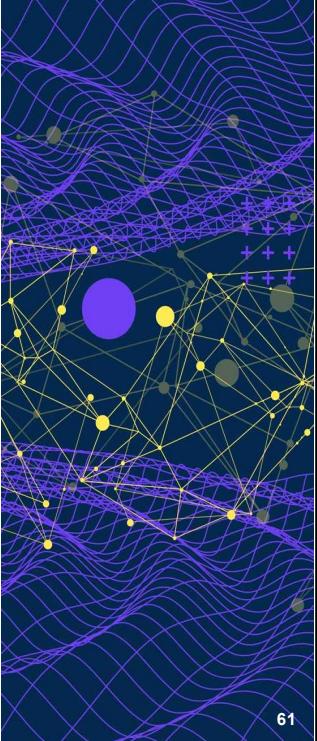
HDS  ISO 27001  ISO 27701  NOOC  RGPD

SNDS  SNO  Spatio  VIDS  Web Information System

Attribution d'un identifiant pérenne

Hébergement des données

Collection/ Sous-ensemble de :



61

Inist-CNRS. Cat OPIDoR. Page entrepôt de données. <https://cat.opidor.fr/>

Cat OPIDoR recense et décrit les services français dédiés aux données scientifiques.

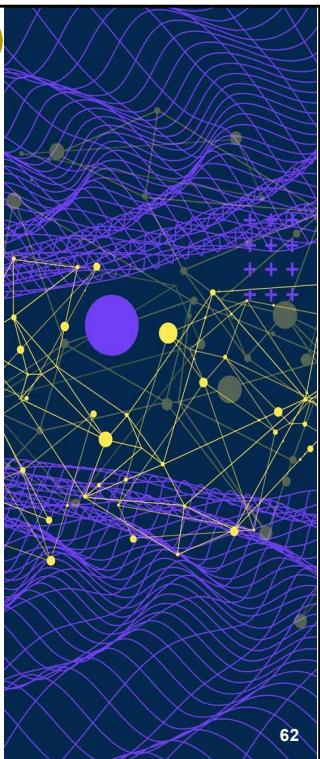
Proposé sous forme d'un wiki, cet outil collaboratif ouvert à tous permet de repérer et ajouter des services pour accompagner les chercheurs sur l'ensemble des étapes clés de la gestion, collecte, stockage, conservation et ouverture des données.

## Recherche d'entrepôts

Repository details	
<b>4TU.Centre for Research Data</b>	
General	Institutions
Name of repository	<b>4TU Centre for Research Data</b>
Additional name(s)	4TU ResearchData formerly: STU DC formerly: STU Dataverse
Repository URL	<a href="https://researchdata.4tu.nl/researchome/">https://researchdata.4tu.nl/researchome/</a>
Subject(s)	<a href="#">Construction Engineering and Architecture</a> <a href="#">Engineering Sciences</a> <a href="#">Natural Sciences</a> <a href="#">Agricultural, Forestry, Horticultural and Veterinary Sciences</a> <a href="#">Chemical Engineering and Technology</a> <a href="#">Earth and Related Environmental Sciences</a> <a href="#">Life Sciences</a> <a href="#">Aerospace, Maritime and Transport Sciences</a> <a href="#">Mathematics and Statistics</a> <a href="#">Social Sciences</a> <a href="#">Computer Engineering and Technology</a> <a href="#">Computer Science</a> <a href="#">Digital and Intelligent Systems</a> , <a href="#">Logistics</a> <a href="#">Systems Engineering</a> <a href="#">Electrical and Computer Engineering</a> <a href="#">Industrial Engineering</a> <a href="#">Manufacturing</a> <a href="#">Materials Science and Engineering</a> <a href="#">Mechanical Engineering</a> <a href="#">Process Engineering</a> <a href="#">Chemical Engineering</a> <a href="#">Metallurgy and Ceramics</a> <a href="#">Mathematics</a> <a href="#">Physics</a> <a href="#">Chemistry</a> <a href="#">Earth Sciences</a> <a href="#">Geography</a> <a href="#">Geosciences</a> <a href="#">Environmental Sciences</a> <a href="#">Ecology</a> <a href="#">Conservancy, Monitoring and Coordinating Activities</a> <a href="#">Biology</a> <a href="#">Botany</a> <a href="#">Zoology</a> <a href="#">Microbiology</a> <a href="#">Microbiology</a> <a href="#">Psychology</a> <a href="#">Cognitive, Behavioral and Evolutionary Sciences</a> <a href="#">Developmental Psychology</a> <a href="#">Educational Sciences</a> <a href="#">Neurosciences</a>
Description	4TU ResearchData, previously known as STU Dataverse, is an archive for research data. It offers the knowledge, experience and the tools to share and safely store scientific research data in a standardized, secure and well-documented manner. 4TU Centre for Research Data provides the infrastructure with advice and support on data management. A long-term archive for scientific research data. Support for current research projects. Tools for reusing research data.
Contact	<a href="https://researchdata.4tu.nl/contact.html#ResearchData@4tu.nl">https://researchdata.4tu.nl/contact.html#ResearchData@4tu.nl</a>

Dans **re3data**, exemple de recherche à partir des critères suivants :

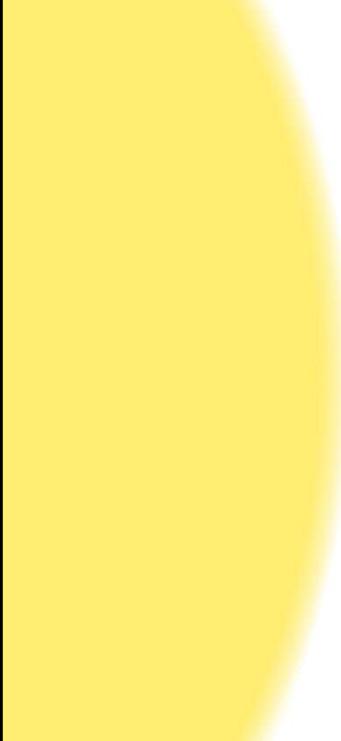
- Informations complémentaires fournies
  - Libre accès aux données
  - Conditions d'utilisation et licence fournis
  - Génération d'un DOI
  - Certification
  - Politique de l'entrepôt fournie



Possibilité de consulter l'annuaire par sujet, type de contenu, pays. La recherche graphique par sujet est matérialisée par une succession de cercles concentriques, du plus général vers le plus spécifique.

Plusieurs entrepôts répondent à ces 6 critères :

- 4TU.Centre for Research Data
  - CLARIN repository at the University of Tübingen
  - NASA Socioeconomic Data and Applications Center
  - PANGAEA



## Identifiants pérennes

## Identifiants pérennes pour les données

- Un jeu de données = un identifiant pérenne (PID)
- Le plus utilisé est le DOI
- Facilite le suivi, la localisation, l'accès et la citation des données lors de leur publication ou à des fins de réutilisation
- Attribué le plus souvent aux données de manière automatique lors du dépôt dans un entrepôt !



Une équipe dédiée de l'Inist-CNRS se tient à disposition pour conseiller dans l'attribution de DOI aux données de recherche et fournit :

- un accès à un espace test pour enregistrer temporairement des DOI et vérifier la compatibilité de ce service avec ses propres workflows
- un préfixe unique de DOI
- un accès à la plateforme DataCite Fabrica pour commencer à créer les DOI
- une assistance à la création et à la conversion de métadonnées...
- un accompagnement dans l'utilisation des différents services proposés par DataCite.

### Ressources :

- *DoRANum. Zoom sur le DOI.* <https://doi.org/10.13143/J5XT-6J41>
- *Inist-CNRS. PID OPIDoR.* <https://opidor.fr/identifier/>

## Identifiants pérennes pour les logiciels

SWHID est l'identifiant pérenne dédié aux logiciels, créé par Software Heritage

Il est possible de déposer les codes sources dans HAL, avec transfert vers Software Heritage



65

Créé par Software Heritage, SWHID (Software Hash Identifier) est un système d'identifiant pérenne sous forme de code alphanumérique qui identifie de manière unique les codes sources des logiciels.

### Ressources :

- DoRANum. Zoom sur SWHID. <https://doi.org/10.13143/3QQG-YX41>
- Software Heritage. <https://www.softwareheritage.org/?lang=fr>
- HAL Documentation. Déposer le code source d'un logiciel.  
<https://doc.archives-ouvertes.fr/deposer/deposer-le-code-source/>

**Source de l'image :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)



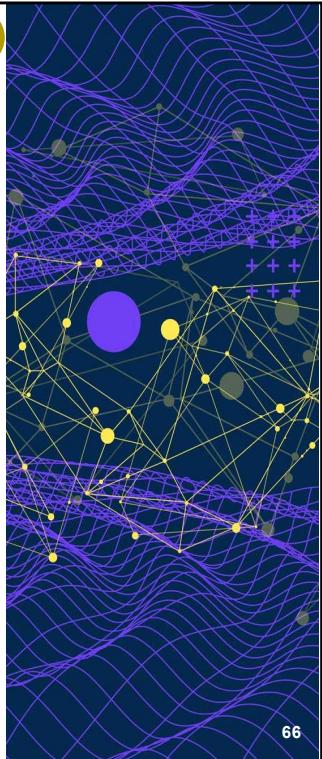
## Identifiants pérennes pour les auteurs et les institutions

Attribuer un **identifiant auteur** (ORCID)

- Fait le lien avec ses productions scientifiques
- Permet d'être bien identifié et cité
- Augmente la visibilité internationale



Pensez à vérifier si votre institution bénéfice d'un **identifiant pérenne ROR**

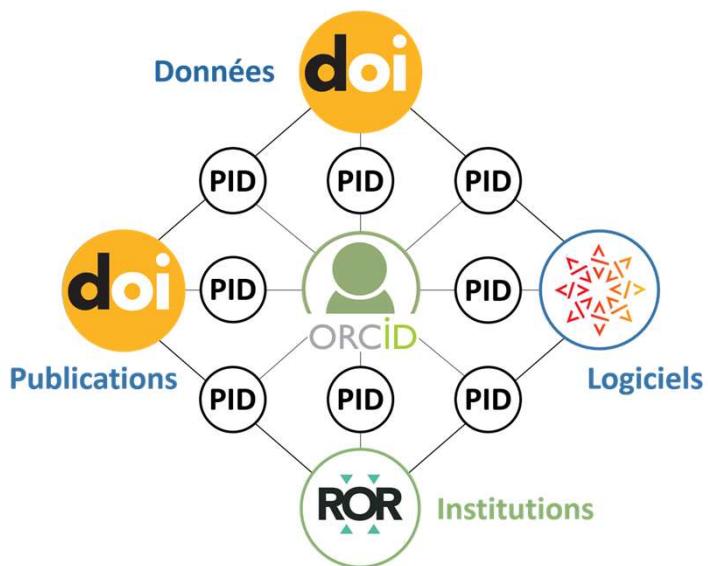


### Ressources :

*DoRANum. Zoom sur ORCID.* <https://doi.org/10.13143/C6RX-9W77>

*DoRANum. Zoom sur ROR.* <https://doi.org/10.13143/H5ZY-BN73>

## Identifiants pérennes – Pour résumer



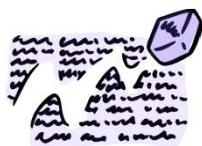
**Source de l'image :** Coralie Wysoczynski – DoRANUM (Inist-CNRS) (CC-BY 4.0)

## Modalités de protection

## Accès et licences



- Accès
  - Dispositif d'accès contrôlé : mot de passe...
  - Accès limité dans le temps par un embargo
  - Accès limité à certaines personnes

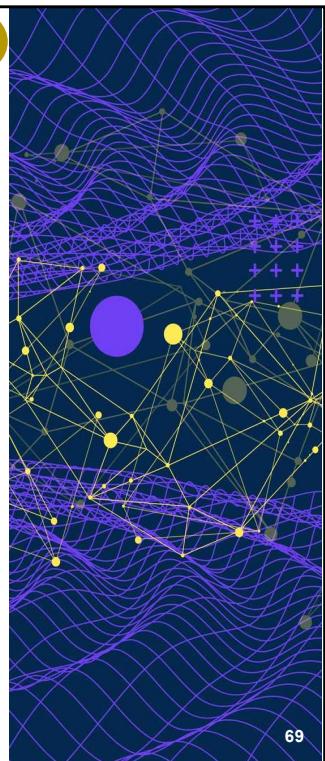


- Chiffrement des données sensibles pour éviter des intrusions malveillantes



- Licences pour éviter une mauvaise (ré)utilisation des données par autrui

CNRS Inist



69

### Accès :

- Accès limité dans le temps par un embargo
  - Pour disposer du temps nécessaire au dépôt de brevets...
  - En fonction de la discipline
  - Peut être déterminé par les éditeurs
- Accès limité à certaines personnes
  - Ex : limitation aux membres du consortium ou à une communauté scientifique

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Licences de diffusion

Attribuer une licence de diffusion à chaque jeu de données permet d'afficher clairement les modalités de réutilisation

- En France → **Licence Ouverte / Open License** (équivalente de la CC-BY) : l'auteur doit obligatoirement être cité
- Licences Creative Commons (CC) avec 4 éléments à combiner :
  - BY : citation de l'auteur
  - NC : pas d'utilisation commerciale
  - SA : partage dans les mêmes conditions
  - ND : pas de modifications
- Licences spécifiques pour les logiciels et les bases de données



**En France**, la licence la plus adaptée pour les données publiques est la Licence Ouverte / Open License Etalab (équivalente de la CC-BY) : l'auteur doit obligatoirement être cité

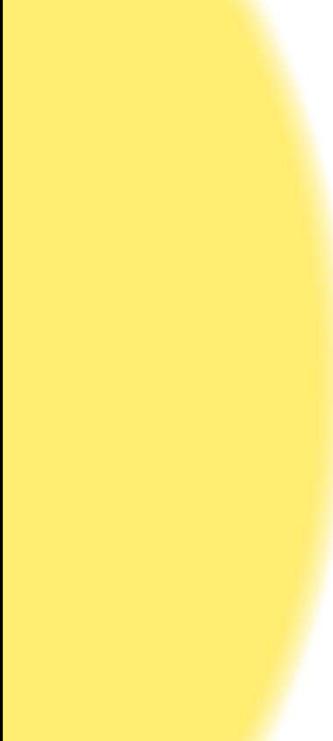
Sinon, il est possible d'utiliser les licences Creative Commons (CC), avec 4 éléments à combiner :

- BY : citation de l'auteur
- NC : pas d'utilisation commerciale
- SA : partage dans les mêmes conditions
- ND : pas de modifications

Il existe des licences spécifiques pour les logiciels et les bases de données.

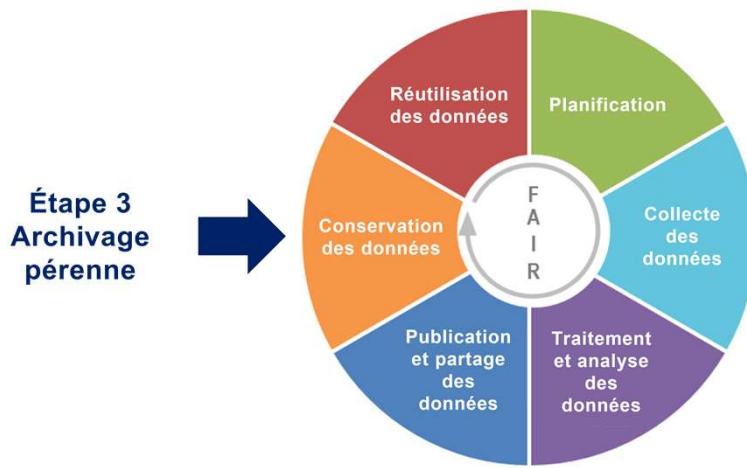
### Ressources :

- *DoRANum. Guide des licences ouvertes.* <https://doi.org/10.13143/TV6F-SV31>
- *DoRANum. Les licences de réutilisation dans le cadre de l'open data et de la loi pour une République numérique.* <https://doi.org/10.13143/SSH2-ZD93>
- *Data.gouv.fr. Licences de réutilisation.* <https://www.data.gouv.fr/fr/licences>
- *Licentia by Inria. License your Data.* <http://licentia.inria.fr/>.
- *License Selector.* <http://ufal.github.io/public-license-selector/>
- *Choose an open source license :* <https://choosealicense.com/>



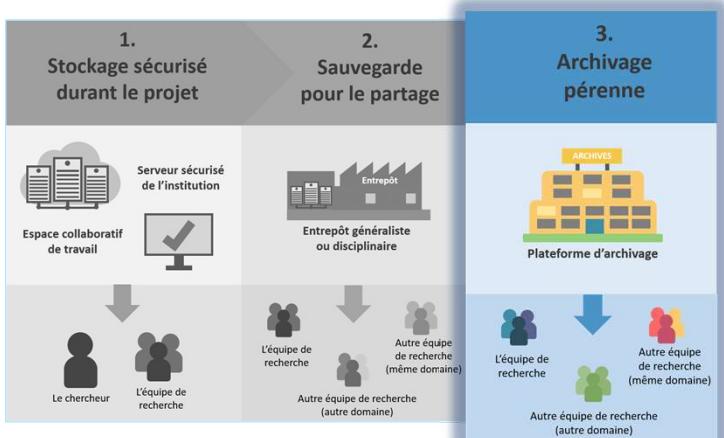
## **Étape 5 : Archivage pérenne**

## Place dans le cycle de vie des données



L'archivage pérenne intervient à la fin du cycle de vie des données même s'il est déjà pensé, réfléchi et préparé bien en amont.

## Archivage pérenne



L'archivage pérenne a pour objectif de conserver les données, d'en garantir l'accès et d'en préserver l'intelligibilité sur le long terme. Comme évoqué lors des journées RIP Data en octobre 2024\*, le long terme est une période difficile à définir. C'est une période de temps suffisamment longue pour que des changements technologiques majeurs surviennent et mettent en péril la lisibilité des formats de fichiers.

\* Journées dédiées à l'archivage, la sélection, la conservation et la suppression des données de recherche, organisées par le groupe de travail Données de la Recherche de la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (DoReMITI)

### Source :

Christine Hadrossek. Synthèse des journées d'étude RIP Data : Quelle sélection, conservation et suppression des données de recherche ?. 2024.

<https://hal.science/hal-04819987v1>

### Ressource :

DoRANum. Stockage, partage et archivage : quelles différences ?

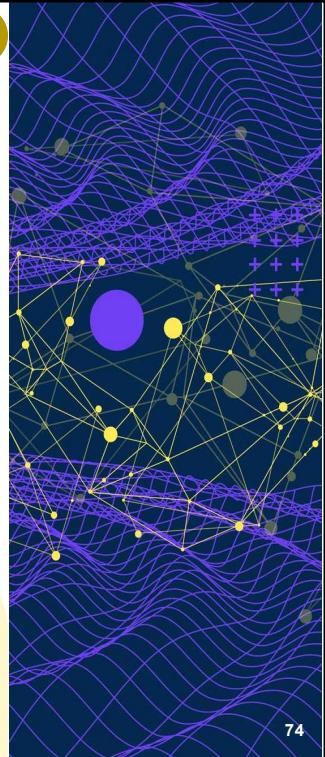
<https://doi.org/10.13143/5DAX-QP58>

## Objectifs :

- **Conserver** les données (supports de stockage)
- Les rendre **accessibles** (formats de fichiers lisibles)
- Préserver leur **intelligibilité** (documentation, métadonnées)



Travail à mener à l'échelle de l'institution,  
en lien avec un archiviste



L'objectif de l'archivage pérenne est de :

- Conserver les données en préservant l'intégrité des supports de stockage ;
- Les rendre accessibles en maintenant la lisibilité des formats de fichiers ;
- Préserver leur intelligibilité afin que les données restent compréhensibles par de potentiels utilisateurs à travers le temps (notamment grâce à des métadonnées spécifiques).

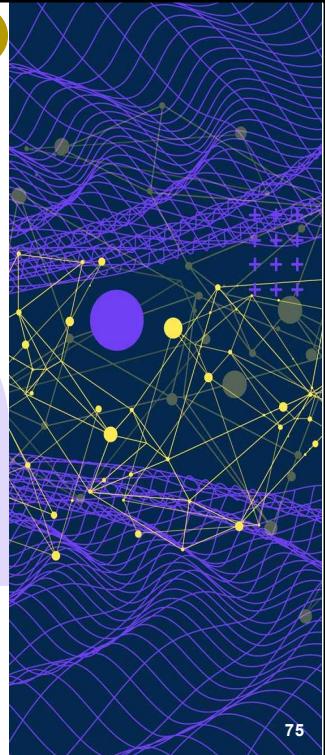
Attention, la responsabilité des décisions d'archivage n'incombe pas uniquement à l'équipe de recherche. Elle doit se faire à l'échelle du laboratoire ou de l'institution en collaboration avec un archiviste.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## Critères de tri :

Déterminer les données à conserver ou à détruire selon ...

- Valeur scientifique des données
- Qualité des données
- Politiques institutionnelles
- Considérations juridiques
- Considérations financières
- Règles de tri et de conservation des archives



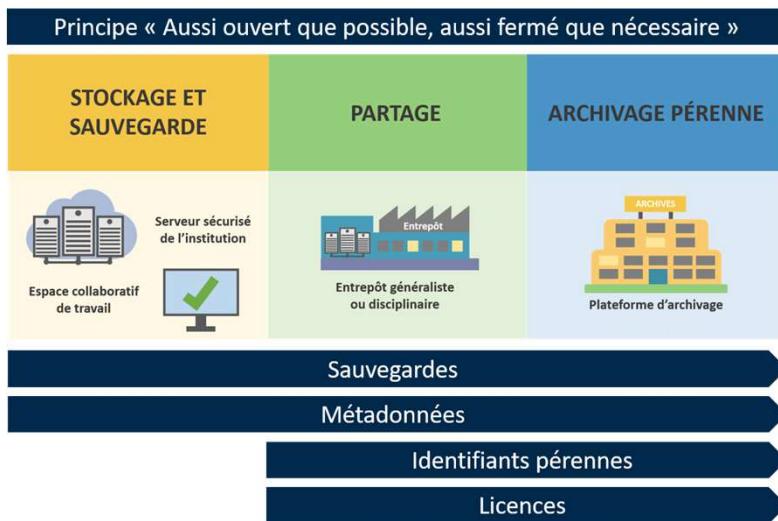
Attention, toutes les données de recherche n'ont pas vocation à être archivées de manière pérenne.

Pour déterminer quelles données doivent être détruites ou conservées à long terme, plusieurs éléments doivent être mis en balance :

- Valeur scientifique des données ;
- Qualité des données ;
- Politiques institutionnelles ;
- Considérations juridiques ;
- Considérations financières ;
- Règles de tri et de conservation des archives

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

## En résumé



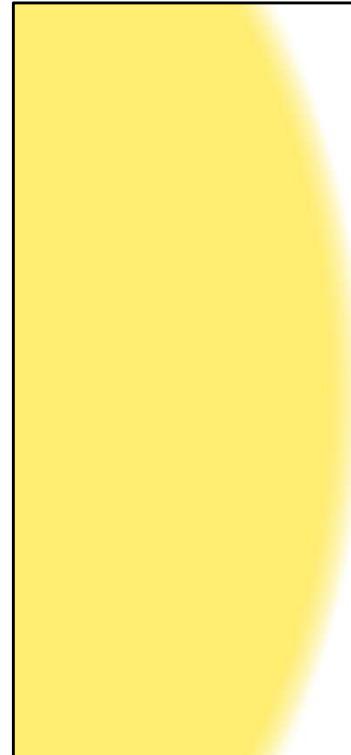
- Le stockage et la sauvegarde sécurisés des données se font durant le projet.  
L'objectif est de garantir la sécurité des données et d'en faciliter l'accès pour l'ensemble des collaborateurs du projet.
- Le partage consiste à déposer les données dans un entrepôt de données afin de les rendre accessibles facilement et de permettre leur réutilisation par des chercheurs du même domaine ou d'un autre domaine, selon les principes FAIR et sur le court et le moyen terme. Le partage se fait souvent à l'issue du projet.
- L'archivage pérenne a pour objectif de conserver les données, d'en garantir l'accès et d'en préserver l'intelligibilité sur le long terme.

Les bonnes pratiques de sauvegarde, d'ajout de métadonnées, attribution d'identifiants pérennes et de licences participent tous au respect des principes FAIR.

### Ressource :

*DoRANum. Stockage, partage et archivage : quelles différences ?*

<https://doi.org/10.13143/5DAX-QP58>



## **Étape 6 : Réutilisation et valorisation des données**

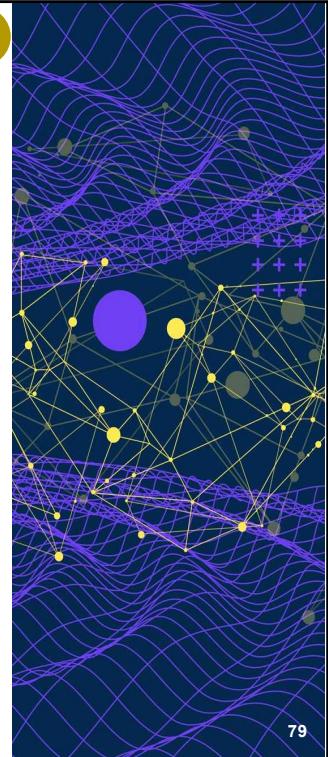
## Place dans le cycle de vie des données



Cette étape intervient à la fin d'un cycle... ou au début !

## Du côté du chercheur

- Plus de visibilité et meilleure valorisation de ses travaux
- Mieux retrouver ses propres données
- Conformité aux demandes des financeurs
- Meilleure accessibilité de la science aux citoyens



Qu'apportent le partage et la valorisation des données pour le chercheur ?

- Plus de visibilité et meilleure valorisation de ses travaux : si des publications sont reliées aux données et que ces dernières sont accessibles, cela renforce la confiance en ces travaux et leur apporte davantage de crédit. Cela permet aussi d'éviter de dupliquer des données ou des résultats déjà publiés.
- Des données mal documentées ou mal conservées sont des données perdues ! Les principes FAIR facilitent la bonne gestion des données, pendant et après le travail de recherche. L'objectif d'un travail de recherche est d'être lu, vérifié, cité, réutilisé. Ceci est impossible si personne n'a accès aux informations, si le format choisi pour le stockage des données n'est plus lisible, si l'information est perdue, ou si l'adresse url d'accès n'est plus disponible !
- Science citoyenne : les publications, les données gagnent à être mise à la disposition de tous, mais pour ce faire, il faut qu'elles soient compréhensibles et manipulables, et accompagnées de métadonnées.

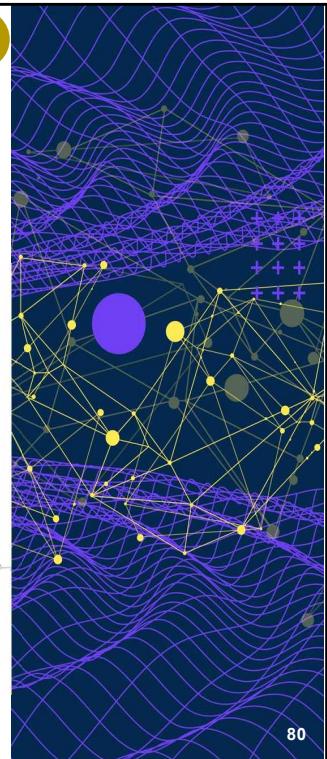
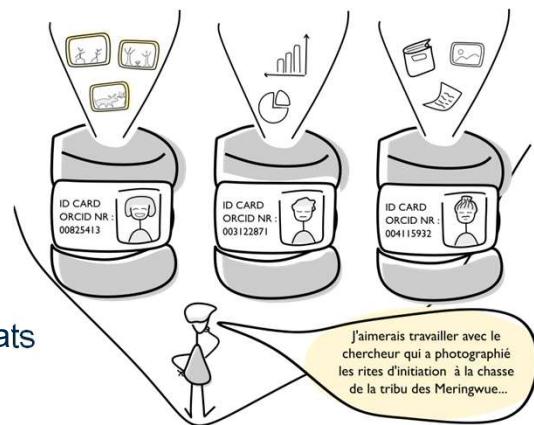
**Source :**

Bibliothèque de l'Observatoire de Paris. GT Science ouverte. Des principes ? Pour quoi FAIR ? 13 avril 2021. <https://www.observatoiredeparis.psl.eu/IMG/pdf/pour-quoi-fair-3.pdf>

**Image :** Lynda Kehli – Inist-CNRS (DoRANum) [CC-BY 4.0]

## Du côté du réutilisateur

- Trouver des jeux de données déjà existants (dans les entrepôts et data papers)
- Découverte de potentiels partenaires
- Reproductibilité et réplicabilité des résultats



Qu'apportent le partage et la valorisation des données au réutilisateur ?

Les chercheurs peuvent trouver des jeux de données déjà existants dans les entrepôts et data papers. De nouveaux partenariats peuvent être envisagés et noués dans le cadre d'une réutilisation des données de la recherche. Enfin, cela permet la reproductibilité et réplicabilité des résultats.

La reproductibilité est la capacité, par une équipe différente, de reproduire une expérience, sans se fier au dispositif expérimental et aux codes logiciels développés par l'équipe d'origine.

La réplicabilité est la capacité, par une équipe différente, de reproduire une expérience en ré-utilisant le même dispositif expérimental décrit (y compris les codes logiciels).

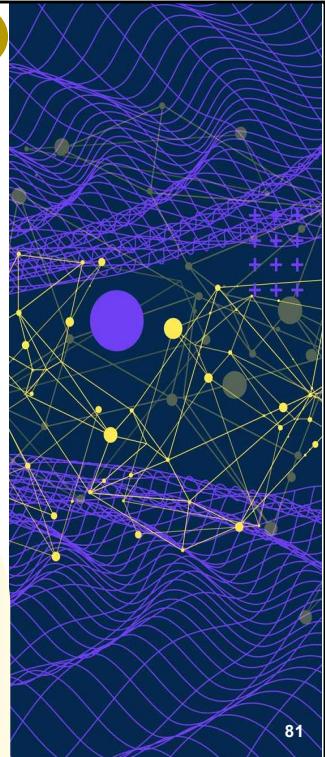
**Ressources pour rechercher des jeux de données :**

- DataCite Commons. <https://commons.datacite.org/>
- OpenAIRE EXPLORE. <https://explore.openaire.eu/search/find/datasets>
- Google Dataset Search. <https://datasetsearch.research.google.com/>

**Image :** Lynda Kehli – Inist-CNRS (DoRANum) [CC-BY 4.0]

## Data journal – data paper

- **Data paper** = article scientifique à part entière
- Décrit des jeux de données de recherche et les métadonnées associées
- 2 possibilités de publication :
  - dans un **data journal** (revue dédiée à ce type de publication)
  - dans une revue classique



Un data paper décrit des jeux de données de recherche et les métadonnées associées. C'est un article scientifique à part entière, suivant le même processus éditorial que les articles scientifiques classiques. Il existe deux possibilités de publication d'un data paper : dans un data journal (revue dédiée à ce type de publication) dans une revue classique.

La structure d'un data paper est composée :

- D'une partie descriptive : éléments communs aux articles classiques (titre, résumé, mots-clés...) + éléments spécifiques aux données (types de données, formats, processus et méthodes de production, métadonnées, réutilisation...)
- D'une partie dédiée à l'accès aux données : identifiant pérenne de données (exemple DOI) permettant d'établir le lien du data paper vers les données préalablement déposées dans un entrepôt.

**Ressources :**

- *Exemples de data paper :*

Richardson Andrew D., Hufkens Koen, Milliman Tom et al. *Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery.*

<https://www.nature.com/articles/sdata201828>

Richardson Andrew D., Hufkens Koen, Milliman Tom et al. *Tracking vegetation*

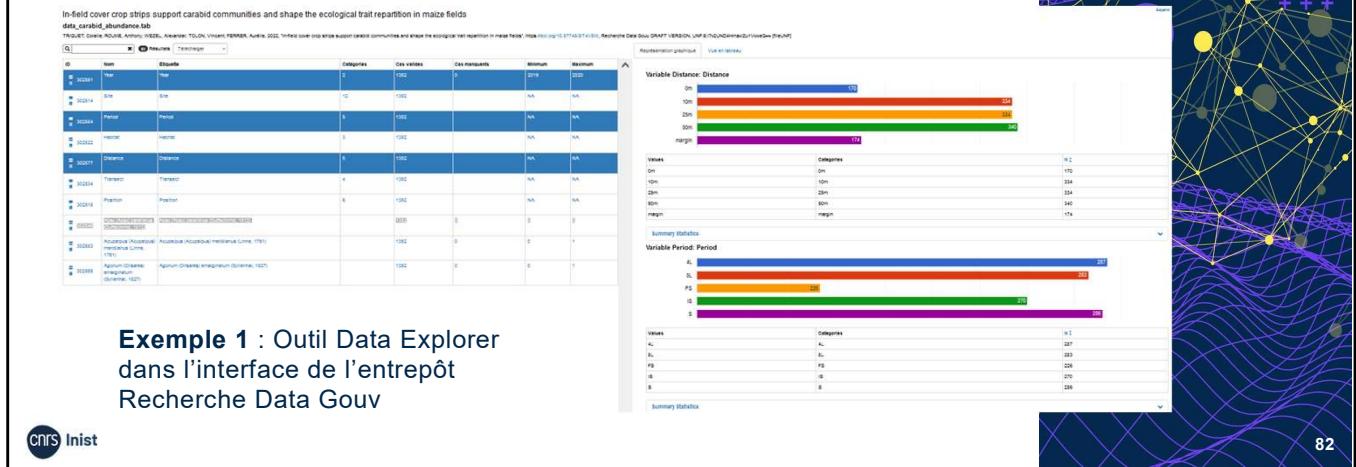
*phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset.* <https://www.nature.com/articles/s41597-019-0229-9>

- *Exemple de data journal : Journal of Physical and Chemical Reference Data.* <https://pubs.aip.org/aip/jpr>
- *Exemples de revues publiant des data papers : Cirad. CoopIST. Publier un Data paper.* <https://coop-ist.cirad.fr/gerer-des-donnees/publier-un-data-paper/4-choisir-la-revue>
- *DoRANum. Thématique data papers et data journals.*  
<https://doranum.fr/data-paper-data-journal/>

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

# Explorer par la visualisation des données

En complément, penser à exposer ses données **sous forme visuelle** (cartographies, graphiques, etc.) via une plateforme



La visualisation devient une nécessité en présence de grandes quantités de données complexes et difficilement appréhendables.

Dans l'interface de l'entrepôt Recherche Data Gouv, l'outil Data Explorer permet d'explorer les données tabulées sous forme de représentation graphique ou en tableau.

L'outil Data Explorer est visible dans « Modalités d'accès au fichier ». Exemple : <https://doi.org/10.57745/DRUUWF>

**Ressource :**

*DoRANum. La visualisation des données : quelques exemples.*

<https://doi.org/10.13143/CSJE-A859>

## Exposition et visualisation des données

L'outil Omeka est un logiciel libre couramment utilisé pour l'exposition et la visualisation des données



**Exemple 2 : site web réalisé avec Omeka**

The screenshot shows a search interface for the CoReA website. The search bar contains "Recherche avancée". Below it, there are filters for "Collection", "Créateur", "Peuples", "Anthroponymes", "Chronologie", "Sujets", "Lieu", and "Type de support". The main search results page displays 770 results for "DÉDICACE DE CORNELIA GRATILLA À LA BONA DEA". It includes a thumbnail of the artifact, its title, author (Jacques Gascou), and a brief description: "Autel de calcaire avec base et couronnement. Pulvini et focus au sommet." Below this, another result is shown: "INSCRIPTION FRAGMENTAIRE SUR PLAQUE DE MARBRE" by Jacques Gascou, with a similar brief description.

On the right side of the slide, there is a large, abstract network visualization on a dark blue background. The network consists of numerous purple nodes of varying sizes and a dense web of purple lines connecting them, representing data connections or relationships.

Exemple : Bibliothèque numérique CoReA (Corpus et Ressources Archéologiques) pour la documentation archéologique du Centre Camille Julian, réalisée avec Omeka. Permet de naviguer dans des corpus et ressources en archéologie. <http://ccj-corea.cnrs.fr/>

Omeka est un outil qui s'inscrit pleinement dans la science ouverte et qui permet la création de bases de données au plus près des principes FAIR.

À partir de données de recherche brutes, l'outil permet de créer des bases de données éditorialisées, autrement dit structurées, accessibles, et visibles sur le web. L'outil offre une grande modularité des fonctionnalités grâce à de nombreux plugins, et traite les divers objets multimédia (textes, images, sons, vidéos).

L'outil offre plusieurs avantages techniques :

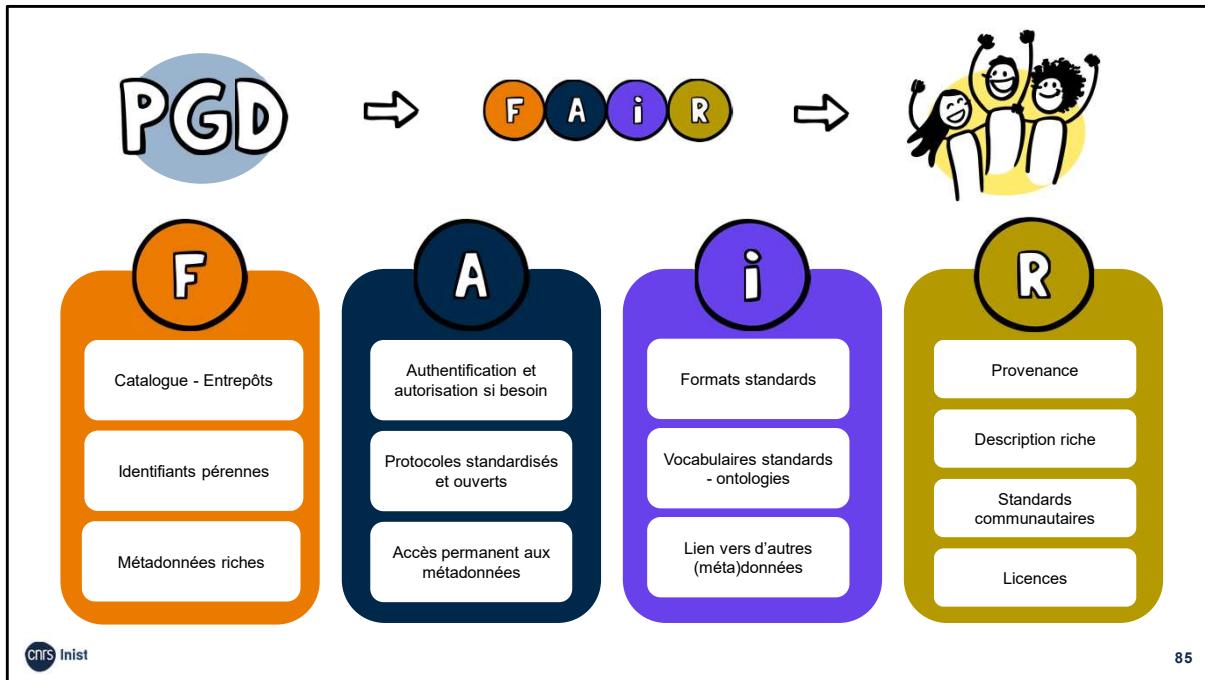
- l'interface est simple et intuitive ;
- les métadonnées sont moissonnables, permettant notamment le référencement dans d'autres bases ;
- une base Omeka peut être connectée à d'autres services grâce à une API REST.

### Ressources :

DoRANum. Thématique accès et visualisation. <https://doranum.fr/acces-visualisation/>

04

## À retenir



En résumé,

- Vous rendrez vos données **Faciles à trouver** en les déposant dans un entrepôt de confiance, vérifiant qu'un identifiant pérenne leur est attribué et en les décrivant précisément avec des métadonnées.
- Vous rendrez vos données **Accessibles** en déterminant qui peut y accéder et comment. Ne pas oublier que si vos données ne sont pas ouvertes, les métadonnées doivent l'être.
- Vous rendrez vos données **Interopérables** en utilisant des standards de votre discipline et utilisant des formats ouverts.
- Vous rendrez vos données **Réutilisables** en documentant vos données de manière à ce que les autres utilisateurs puissent les comprendre et en leur attribuant une licence.

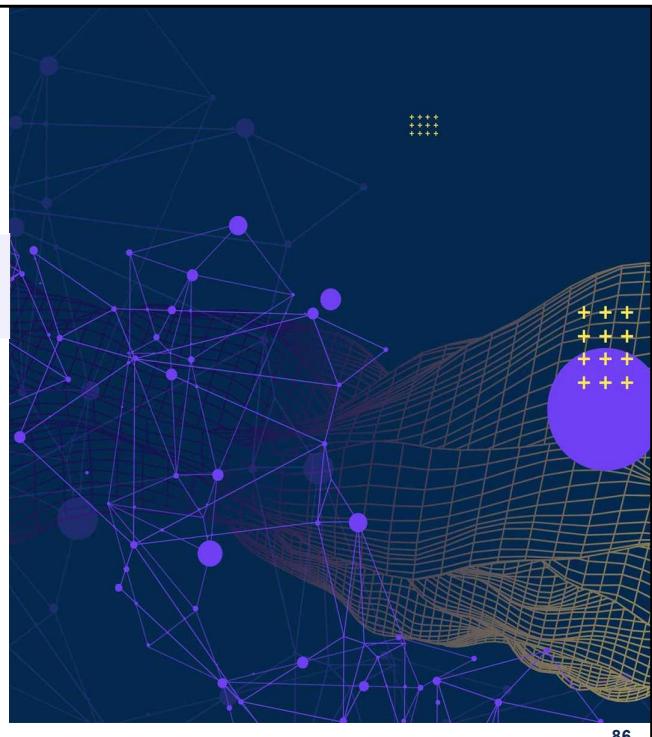
Si vous rédigez votre **Plan de Gestion de Données** en pensant aux bonnes pratiques à mettre en œuvre en suivant les principes FAIR, vous atteindrez une gestion optimale de vos données.

**Source des images :** Coco Material (<https://cocomaterial.com/>) (CC0 1.0)

Merci de votre attention

contact@doranum.fr

Pour en savoir plus, rendez-vous sur  
<https://doranum.fr>



86