

# Gestion et ouverture des données de la recherche

La **science ouverte** vise à généraliser l'accès sans entrave aux résultats d'une recherche plus collaborative, transparente, **intègre** et de qualité. Dans ce contexte, les organismes, politiques et financeurs se mobilisent pour que les données de la recherche soient gérées dans l'esprit **FAIR** en vue de leur préservation et réutilisation, et si possible partagées selon le principe « aussi ouvertes que possible, aussi fermées que nécessaire ».



EPRIST

## Bonnes pratiques de gestion



Les **principes FAIR** définissent des bonnes pratiques pour faciliter la découverte, l'accès, l'interopérabilité et la réutilisation d'un jeu de données, par les êtres humains et les machines, dont notamment :

- Choisir des **formats de fichier ouverts**, standards
- Décrire ses données à l'aide de **métadonnées riches**, et selon des standards
- Déposer ses données dans un **entrepôt de confiance**
- Attribuer des **identifiants uniques et pérennes** aux données (DOI par ex.)
- Attribuer une **licence de diffusion et réutilisation** (licence ouverte, Creative Commons...)
- Indiquer clairement la provenance (origine et historique) des données
- Utiliser des **vocabulaires contrôlés** (exemples)
- Garantir l'accès ouvert aux métadonnées même si les données sont fermées ou en accès restreint

## Quelles données ?

Les données sont des informations numériques de divers types (observation, expérimentation, simulation...) et natures (chiffres, textes, images, sons...), collectées ou produites dans le cadre d'un projet de recherche ou d'activités scientifiques. Elles sont généralement regroupées dans un ensemble homogène et/ou cohérent appelé jeu de données (dataset).  
Voir comment **trouver des jeux de données**, et les **citer**.

### DONNÉES PUBLIÉES

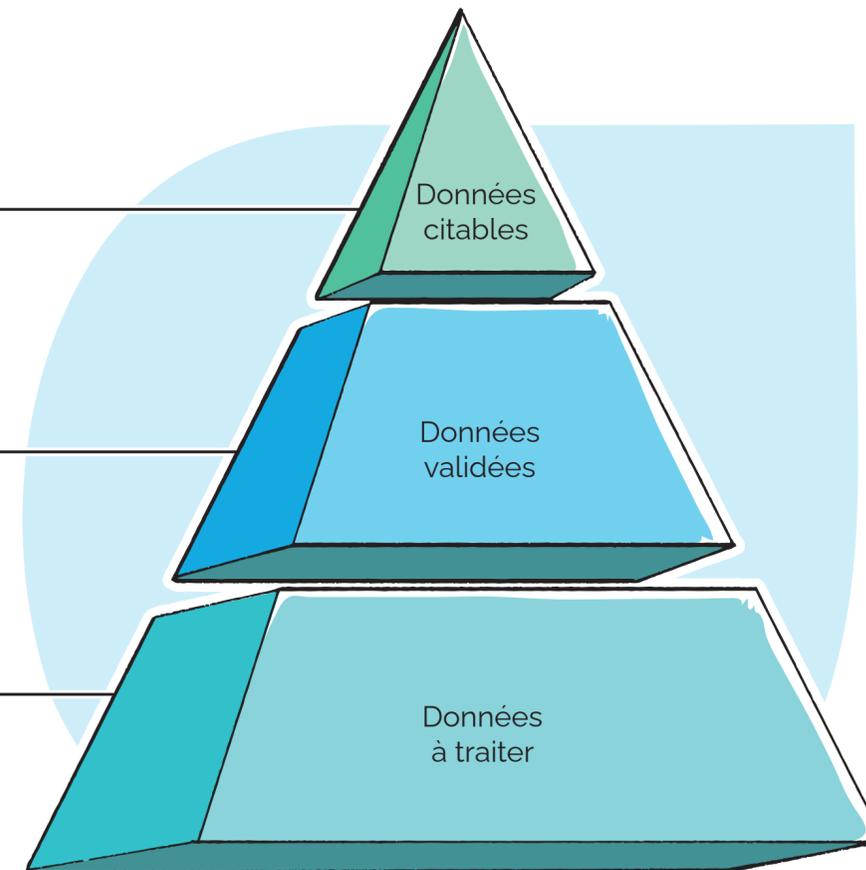
*dans un entrepôt*

### DONNÉES CURÉES

*(enrichies, corrigées, structurées...)*

### DONNÉES BRUTES

*(données d'observation, d'expérimentation, de simulation...)*



Bien gérer ses données est essentiel pour faciliter leur exploitation, sécurisation et préservation tout au long de leur **cycle de vie**, et le cas échéant faciliter leur **ouverture** au public et leur **réutilisation**.  
Pour cela, il est recommandé de :

## 1- Documenter

Au démarrage d'un projet de recherche, rédiger un **plan de gestion des données** (PGD) pour expliciter la façon dont les données produites seront obtenues, décrites, documentées, traitées, stockées et disséminées. Ce document précise notamment les responsabilités, ainsi que les **coûts et ressources** nécessaires à la gestion et au partage des données. Il doit être régulièrement mis à jour au cours du projet.



Voir le **webinaire sur la rédaction d'un PGD**, et le **tutoriel** sur l'outil d'aide à la rédaction DMP OPIDoR. De nombreux modèles de PGD sont disponibles (**exemples**).

Certains établissements d'enseignement supérieur et de la recherche offrent un service d'accompagnement à la rédaction des PGD, consulter **SOS-PGD**.



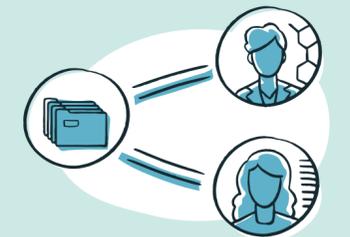
## 2- Conserver

Durant le projet, **stocker et sauvegarder** les données sur une infrastructure adaptée de l'institution afin de les sécuriser et en faciliter l'accès pour les collaborateurs du projet. Pour faciliter le repérage et la compréhension des données, adopter des bonnes pratiques **d'organisation et de nommage** des données.

A la fin du projet, assurer un **archivage pérenne** des données sur une plateforme d'archivage ou un entrepôt pour en garantir l'accès sur le long terme (+ 30 ans), en privilégiant des formats de fichier ouverts, standards ou largement utilisés (**exemples**).

## 3- Partager

Le **dépôt dans un entrepôt** s'inscrit dans une démarche de partage encouragée, voire exigée par les politiques actuelles, que les données soient diffusées en accès restreint (accessibles sur demande, sur justification...) ou ouvertes à tous et librement réutilisables.



Il existe de nombreux entrepôts, de **plusieurs types** : **Institutionnels**, national (Recherche Data Gov), disciplinaires, multidisciplinaires... Ils sont répertoriés dans **FAIRsharing** et **Re3data**. Certains entrepôts permettent de contrôler l'accès aux données. Voir comment **choisir un entrepôt**.



*Le partage des données issues des projets financés sur fonds publics doit se faire en respectant les **règles juridiques, éthiques** et contractuelles applicables à ces données. Voir les outils d'aide à la décision.*