

# Fiche synthèse du chapitre Je découvre les données de la recherche

## Module doctorant B10 : Comment bien gérer ses données pour faciliter son travail de doctorat ? Enjeux et bonnes pratiques

### Définition

« Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme **sources principales pour la recherche** scientifique et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats de recherche** » ([OCDE](#), 2007).

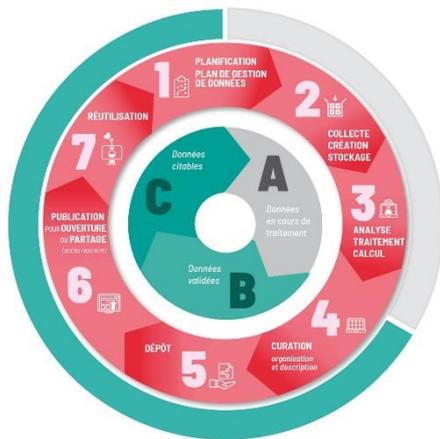
### Types de données

On peut distinguer 5 grands types de données :

- D'observation, capturées en temps réel (exemples : données d'enquête, données d'imagerie médicale, données de capteurs),
- Expérimentales, obtenues à partir d'équipements de laboratoire (exemples : tests de résistance de matériaux, séquençage génomique, chromatogrammes),
- De simulation, produites par des modèles expérimentaux (exemples : des modèles économiques, simulations sismiques, et climatologiques),
- Dérivées ou compilées, issues du traitement ou de la combinaison de données "brutes" (exemples : des bases de données compilées),
- De référence, par exemple des corpus, généralement publiés et édités, de contenus faisant référence dans un domaine (exemples : bases de données de gènes, collections d'archives).

Ces données peuvent prendre différentes formes : documents (texte), tableurs, photographies, films, transcriptions, échantillons, séquences génomiques.

## Cycle de vie des données



Les données évoluent au fil de la recherche, le cycle de leur vie va de leur collecte jusqu'à leur archivage.

Les données de la recherche suivent plusieurs étapes, depuis la collecte ou la création, jusqu'à l'archivage en passant par le traitement, la description et la diffusion. Ces étapes constituent le cycle de vie des données.

## Pourquoi bien gérer ses données ?

Pour être efficace dans sa recherche, il est important d'anticiper et de prendre de bonnes habitudes dès le début de votre thèse et d'avoir conscience des bonnes pratiques à mettre en œuvre à toutes les étapes de la vie des données.

Une bonne gestion des données, c'est pour soi, ses collègues, son jury de soutenance, la communauté de recherche.

Bien gérer ses données facilite le travail de recherche (évite une perte de temps, une perte de données, une perte de notes, des oublis, etc.) et permet ensuite de valoriser ses données.

Enfin, bien gérer ses données c'est aussi respecter la transparence et l'éthique scientifique.

## Que trouve-t-on dans un jeu de données bien construit ?

Pour comprendre un jeu de données, il est nécessaire que leur description soit détaillée et respecte les principes [FAIR](#) (Facile à trouver, Accessible, Interopérable, Réutilisable).

Quelques éléments importants : titre, auteur, date, descriptif, mots-clés/taxonomie, couverture, format, type, volume, paramètres, contexte de productions, read-me, financeurs, licence d'usage...

## Où trouver des données ?

Il est recommandé, dans son état de l'art, de prendre en compte les données de la recherche au même titre que les publications.

### Les données publiques

Les données publiques sont mises à disposition des citoyens via des plateformes comme [data.gouv](https://data.gouv.fr) (pour la France) et [data.europa.eu](https://data.europa.eu) (pour l'Europe).

### Les données de la recherche

Les données de la recherche sont accessibles :

- en association avec des publications
- sur des entrepôts de données, généralistes ou nationaux, comme Recherche Data Gouv ou thématiques, comme Progedo (SHS) et Pangaea (Sciences de l'environnement).

Dans chaque entrepôt, il est possible de trouver des jeux de données.

Il existe plusieurs répertoires d'entrepôts :

- la liste évolutive des [entrepôts](#) sur Recherche data gov
- [Cat Opidor](#)
- [Re3data](#)

Pour trouver des données, il est également possible d'utiliser des moteurs de recherche spécialisés :

- [OpenAIRE Explore](#) (jeux de données européens)
- [DataCite Commons](#)
- [Google Datasetsearch](#)

## Citer un jeu de données de la recherche

Comme toutes les autres ressources utilisées, il est nécessaire de citer correctement un jeu de données dans la thèse.

La plupart des entrepôts propose, pour chaque jeu de données, une rubrique « How to cite ». A partir du doi, l'outil [DOI Citation Formatter](#) permet de générer la citation selon différents styles.

Il est possible, avec Zotero, d'intégrer des jeux de données dans sa bibliographie.

## Qui peut vous aider ?

La Cellule Data Grenoble Alpes, vous accompagne tout au long de votre thèse et du cycle de vie des données, depuis la collecte jusqu'à la diffusion en passant par le stockage, la description, etc. Les services de la cellule sont détaillés sur le site [Science Ouverte de l'UGA](#).

Adresse de contact (par exemple pour des rendez-vous personnalisés) : [sos-data@univ-grenoble-alpes.fr](mailto:sos-data@univ-grenoble-alpes.fr)

Vous pouvez également contacter le [référent données](#) de votre laboratoire.

## Ressources complémentaires

Le site web [science ouverte UGA](#) propose de nombreuses ressources sur la gestion des données mais aussi sur les publications, les codes et logiciels de recherche, ainsi que sur la démarche science ouverte.

Le site d'autoformation [DoraNum](#) est également un réservoir précieux de ressources et de fiches pratiques.

Enfin, le Passeport pour la Science ouverte à l'usage des doctorants, disponible [en français](#) ou [en anglais](#) du site [Ouvrir la Science](#) pourra également vous accompagner durant votre thèse.