

## **14. La gestion des données de la recherche**

**Les données  
de la recherche sont définies  
comme des enregistrements factuels  
(chiffres, textes, images et sons)  
utilisés comme sources principales  
pour la recherche scientifique  
et nécessaires à la validation  
de ses résultats<sup>1</sup>.**

En pratique, le terme recouvre un large éventail de matériaux : séquences de génome, retranscription d'entretien de patients en psychologie, photographies de fouilles archéologiques, tableaux statistiques d'enquêtes en sciences sociales, bases de données textuelles.

L'approche adoptée pour leur gestion variera selon les formats, le travail dont elles sont le produit (données d'observation, expérimentales, de simulation, brutes ou compilées, etc) et des lignes directrices fournies par les bailleurs de fonds.

## Pourquoi doivent-elles être gérées ?

La bonne gestion des données est une part essentielle d'un projet de recherche. Elle contribue à son bon déroulement et permet de gagner du temps en permettant de disposer des informations nécessaires à chaque étape. De nombreuses agences de financement de la recherche demandent de produire un Plan de Gestion de Données/Data Management Plan (PGD/DMP), document qui rend compte des problématiques et des choix en matière de gestion des données produites tout au long du processus. (Guide 11 – Les Plans de Gestion de Données)

De plus, dans le cadre de la Science ouverte, l'attention plus grande portée au partage des données implique de les préparer dans la perspective d'une possible réutilisation par d'autres chercheurs que ceux directement impliqués dans le projet en cours.

## Comment les données sont-elles partagées ?

Les données sont déposées pour être sauvegardées dans des entrepôts de données, généralistes tels *Zenodo* en Europe <https://www.zenodo.org/>, *Dryad* <http://datadryad.org/> aux Etats-Unis, spécialisées par disciplines et formats, ou institutionnels, sur le modèle du logiciel <https://dataverse.org> de Harvard University. Pour les choisir, des annuaires des entrepôts sont consultables : le *Registry of Research Data Repository* <http://www.re3data.org> et pour les entrepôts en France *Cat OPIDoR* <https://cat.opidor.fr/>.

Lorsqu'on partage des données, il est important de penser à la façon dont elles pourront être réutilisées par d'autres. L'application des principes FAIR à la gestion des données constitue une ligne directrice afin de garantir que les données soient :

- Faciles à trouver (Findable) – s'assurer que les données peuvent être repérées à l'aide de leurs métadonnées et d'un DOI.
- Accessibles (Accessible) – privilégier l'usage de logiciels open source et de formats universellement reconnus.
- Interopérables (Interoperable) – utiliser des standards et des formats de données compatibles avec différents systèmes.
- Réutilisables (Reusable) – s'assurer que les données puissent être utilisées et exploitées par d'autres personnes en les dotant d'une licence appropriée comme la Licence Ouverte ou l'Open Database License (ODbL).

Les entrepôts devraient suivre ces lignes directrices autant que possible, notamment pour maintenir l'accessibilité des données. Un bon entrepôt s'engagera à une préservation à long terme du matériel et à l'accès à celui-ci, fournissant un lien permanent afin que les chercheurs n'aient pas à mettre à jour constamment les liens dans leur profil en ligne. Il est toujours préférable de choisir un entrepôt certifié ou en voie de certification par le *Core Trust Seal*<sup>2</sup>. Certains entrepôts facturent des frais de stockage, notamment si le jeu de données est très important. Ce coût doit être anticipé en amont afin d'éviter une mauvaise surprise à la fin du projet.

---

2 - <https://www.coretrustseal.org>

## Principes de la gestion des données de la recherche

### Organisation

Une bonne gestion des données débute par l'organisation de l'information. Les chercheurs ont besoin d'installer un système prenant en compte les informations physiques et numériques et leur permettant de trouver facilement et rapidement ce dont ils ont besoin. Plus l'équipe est large et dispersée et les données nombreuses plus les questions d'organisation des dossiers, de nommage des fichiers et de disponibilité de l'information sont critiques.

### Stockage & Sauvegarde

Il convient de distinguer les besoins de stockage en cours de projet et ceux d'archivage pérenne à son issue. La méthode et l'infrastructure de stockage pendant le projet, la fréquence des mises à jour, le volume et l'infrastructure des sauvegardes pérennes après le projet doivent être anticipées.

### Les données personnelles et/ou sensibles

Contrairement aux publications, l'ouverture en vue d'une réutilisation est le statut juridique par défaut des données publiques y compris produites dans le cadre des activités de recherche (loi Valter de 2015, loi Lemaire de 2016). Toutefois, de nombreuses données sont protégées par un droit spécifique ou une réglementation particulière. La protection des données personnelles requiert par exemple une attention spécifique. Il en est de même avec les données sensibles, relevant de la défense nationale, du secret professionnel, des secrets industriels et commerciaux. Dans ces différents cas, tout en étant avantageusement gérées selon les principes FAIR, les données resteront fermées. Le cadre juridique est complexe. Des guides juridiques permettent une première approche en relation avec les spécificités de chaque projet<sup>3</sup>.

### Partage

Les financeurs encouragent fortement le partage des données selon le principe *aussi ouvert que possible, aussi fermé que nécessaire*, grâce à un entrepôt approprié, en utilisant les descriptions adéquates et une licence adaptée. Si toutes les données doivent être gérées, toutes ne peuvent être partagées.

Librement adapté du guide de l'Office of Scholarly Communication, University of Cambridge, sous licence CC-BY

---

3 - [https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Guide\\_Juridique\\_V2.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Guide_Juridique_V2.pdf)