

15. Les Plans de Gestion de Données

**Les Plans
de Gestion de Données (PGD)
ou *Data Management Plan (DMP)*
sont des documents d'accompagnement
détaillant les choix retenus
en matière de gestion, de partage et
d'archivage des données pendant
et après un projet
de recherche.**

Les agences de financement tendent à les systématiser selon un cadre commun susceptible de correspondre à la grande variété des données, 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 résultats¹.

¹ - Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds Publics. [En ligne] URL : www.oecd.org/sti/sci-tech

Pourquoi sont-ils nécessaires ?

L'ANR – Agence Nationale de la Recherche - rend obligatoire le dépôt d'un PGD/DMP pour les projets déposés en 2019. Ils l'étaient déjà pour les projets européens (ERC/Horizon 2020). Au-delà de la demande d'attester de la conception et de la mise en œuvre des procédures de gestion nécessaires, les PGD s'inscrivent dans les bonnes pratiques académiques. La planification de la gestion des données doit permettre d'anticiper des difficultés éventuelles et des dépenses nécessaires, de perdre moins de temps en cours de recherche et moins de données sur le long terme. Evidemment, tout ne peut être anticipé, c'est pourquoi le PGD se conçoit comme un document évolutif à réviser à intervalles réguliers, précisés par les bailleurs de fonds. L'importance accrue accordée à l'archivage pérenne et au partage des données en fin de projet est un autre facteur à considérer qu'un PGD mené à son terme doit faciliter.

Le contenu d'un Plan de Gestion de Données

La liste suivante comprend les points communs à de nombreux PGD. Chaque bailleur de fonds dispose de ses propres exigences et il est toujours préférable de les vérifier avant de préparer le document². Différents modèles de PGD sont accessibles sur DMP OPIDoR : <https://dmp.opidor.fr/>.

Sources des données

Les chercheurs doivent considérer la source des données qu'ils utiliseront, y compris celles recueillies antérieurement par d'autres personnes. Il importe également de décrire la méthode de recueil ou de constitution des données. Les choix de source et de méthode doivent être justifiés.

Types de données

Une grande partie du PGD consiste à décrire le(s) type(s) de données utilisées et produites au cours de la recherche. Cette information devra inclure des détails sur la forme que prendront les données, les formats des fichiers, les métadonnées utilisées pour décrire les données et les raisons pour lesquelles elles ont été choisies. À ce stade, il peut être nécessaire d'envisager des formats plus appropriés à la conservation, au partage, à l'interopérabilité de long terme que les formats habituellement utilisés. Le type de données collectées aura une influence sur toutes les autres parties du PGD, il est donc important d'y réfléchir attentivement.

Stratégies de sauvegarde

Une fois les données collectées, il est nécessaire de les conserver en sécurité pour le projet en cours et pour leur préservation future, donc dans une double logique de stockage et d'archivage. Les chercheurs auront besoin de préciser les méthodes de sauvegarde et de justifier leur choix et d'anticiper les volumes, les infrastructures et les coûts, à inclure dans les demandes de financement. Ils pourront avoir recours à des entrepôts de données certifiés, dont la liste est accessible sur r3data.org (<https://www.re3data.org/>).

Le partage des données

L'accent accru mis sur le partage des données est un autre point d'attention pour les chercheurs. Qui pourra avoir accès aux données pendant la vie du projet ? Comment cet accès sera-t-il géré ? La protection des données sensibles devra être assurée. Les chercheurs devront également envisager le partage à long terme des données et la manière de l'assurer, en retenant des entrepôts et des formats de données adaptées.

Questions juridiques et éthiques

Composer avec différents types de données requiert d'être attentif aux questions de confidentialité, critiques avec des données personnelles ou sensibles. Les chercheurs devront penser à la façon dont les questions éthiques ou juridiques seront gérées, ce qui inclut le lieu où les données seront stockées, la façon dont elles seront partagées avec l'équipe et le monde entier et les questions de droit d'auteur. La définition des données sensibles est large et varie de pays en pays, les chercheurs devront vérifier les pratiques locales³.

Responsabilité pour les données

Pour éviter la dispersion des responsabilités au sein d'équipes nombreuses et mobiles, les PGD devront identifier un responsable des données, contact principal pour ces questions et responsable de leur sauvegarde à l'issue du projet.

Adapté du guide de Cambridge University (CC-BY) : <https://osc.cam.ac.uk/>

2 - Le modèle ANR est accessible ici : <https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf>.

Pour la France : https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Guide_Juridique_V2.pdf.