



LA GESTION DES DONNÉES

Cette fiche, issue des travaux préliminaires et des consultations menées par l'équipe projet du Programme prioritaire de recherche (PPR) Autonomie, piloté par le CNRS, a pour objectif de présenter un certain nombre de ressources pour faciliter le travail des équipes dans la réutilisation des données de recherche, leur dépôt dans des entrepôts de données et leur diffusion éventuelle. Elle fait partie d'une série de trois fiches pratiques, dont l'une s'intéresse à la diffusion des données et l'autre à la réutilisation des données.

Le plan de gestion des données

Le contexte de la science ouverte, c'est-à-dire l'ambition de rendre la recherche scientifique et les données qu'elle produit accessibles à tous les niveaux de la société. [Le décret n° 2021-1572 du 3 décembre 2021](#) rend désormais incontournable la rédaction d'un PGD¹ dans tout projet de recherche ou dans le cadre d'une thèse.

Le PGD a pour objectif de faciliter les procédures de collecte, stockage, gestion, documentation, pérennisation, partage, diffusion, archivage, tout en s'interrogeant sur les coûts de ces étapes et les enjeux juridiques, éthiques, pratiques et scientifiques. Il ne s'agit pas d'un document figé ; il est voué à évoluer en même temps que la recherche et au fur et à mesure que sont prises de nouvelles décisions, que sont résolues des difficultés, etc.

Le PGD est aussi un outil indispensable pour préparer la réutilisation éventuelle des données et leur partage une fois le projet de recherche terminé. En invitant à considérer les données produites et leur stockage sous l'angle des principes

FAIR², il facilite grandement la ré-exploitation des données, que ce soit par le ou les chercheurs producteurs ou d'autres équipes. Il permet de gagner du temps lors de la phase de mise à disposition en ayant organisé la documentation des méthodes et des jeux de données et en ayant traité les questions juridiques.

BON À SAVOIR

Rédiger un PGD est une obligation formulée par l'ANR ainsi que par de nombreux financeurs, notamment au niveau européen. Dans le cadre d'un projet ANR, une première version doit être remise 6 mois après le démarrage du projet, puis à mi-parcours si le projet dure plus de 30 mois et une fois le projet terminé. Cette obligation peut être renforcée et étendue par votre établissement dans le cadre du décret relatif au respect des exigences de l'intégrité scientifique par les établissements publics. N'hésitez pas à vous rapprocher de votre établissement de tutelle pour voir si des ressources ou des instructions sont déjà mises en place.

1. Plan de gestion de données.

2. Faciles à trouver, accessibles, interopérables, réutilisables.

Pourquoi faire un PGD ?

La rédaction d'un PGD répond à des enjeux juridiques, éthiques, pratiques et scientifiques.

Du côté juridique et éthique

- Connaître la conformité du projet de recherche au RGPD³ et aux normes éthiques qui vous semblent pertinentes : coconstruction et restitution aux personnes enquêtées par exemple ;
- Définir les protocoles à suivre en cas de rupture de la confidentialité ;
- Connaître la propriété intellectuelle des données collectées et réutilisées.

Du côté pratique

- Décrire les procédures de collectes des données, les types de données traitées ;
- Définir la manière dont le stockage va être réalisé afin d'éviter la perte des fichiers ;
- Préciser les standards mobilisés dans la description des données ;
- Mettre à plat la responsabilité de chacun et définir des rôles de référents vis-à-vis des données ou du RGPD ;
- Construire un projet « données » commun et d'anticiper les partages de données internes au consortium.

Du côté scientifique

- Prévoir la manière de documenter les données au fur et à mesure de la recherche et les modes de documentations, les outils nécessaires, les problèmes méthodologiques qui peuvent se poser ;
- Consigner les divers recodages, choix et tris qui peuvent être décidés tout au long du traitement des données. Ces informations sont indispensables au moment du partage des données entre les membres d'une même équipe, d'un consortium, d'un laboratoire ou lors du partage plus large au reste de la communauté scientifique. Cette étape de documentation au cours de la recherche est plus facile et plus rapide que de ne documenter qu'a posteriori les données pour leur diffusion.

3. Règlement Général sur la Protection des Données. Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.

BON À SAVOIR

Toute la documentation produite dans le cadre de la rédaction d'un PGD est valorisable sous la forme de *data paper*, c'est-à-dire une publication consacrée à la valorisation de données et mise à disposition de la communauté de recherche.

Ressources

Être accompagné par des professionnels

[Le répertoire des SOS-PGD⁴](#) recense un grand nombre de services susceptibles de répondre à vos questions et vous accompagner, tels que les services documentaires de nombreuses structures de recherche ou des services répartis sur l'ensemble du territoire comme les MSH⁵.

Les bibliothèques universitaires et centres documentaires peuvent aussi orienter et conseiller sur ces sujets.

Recherche Data Gouv⁶ mets en place des ateliers de la donnée qui accompagnent, tout au long du cycle de vie des données, les équipes de recherches. [Une première liste d'Ateliers labéllisés est disponible.](#)

Pour les chercheurs affiliés au CNRS, [le département « Valoriser les données de la recherche »](#) de l'INIST⁷ peut être mobilisé afin d'attribuer un DOI⁸ aux données avant la fin de leur traitement, même si elles n'ont pas vocation à être partagées.

Les DPO⁹ de chaque institution peuvent répondre aux questions relatives à législation sur les données personnelles.

Des centres de ressources labéllisés

[DORANum¹⁰](#) propose des supports de formation et d'autoformation (tutoriels et webinaires) sur la gestion des données et sur les PGD.

4. Services opérationnels de soutien à la rédaction des PGD.
5. Maison des sciences de l'homme.
6. L'écosystème français au service du partage et de l'ouverture des données de recherche.
7. Institut de l'information scientifique et technique.
8. *Digital Object Identifier*, identifiant pérenne et unique, permet de référencer, citer et fournir un lien stable vers un fichier en ligne.
9. Délégué à la protection des données.
10. Données de la recherche : apprentissage numérique.

[DMP OPIDoR](#), outil développé par OPIDoR¹¹, recense différents modèles de PGD propres à des structures ou à des financeurs, ce qui permet de s'inspirer ou d'adapter des PGD existants. Il est à noter que l'ANR y propose un modèle. OPIDoR assiste aussi les équipes qui désireraient décliner cet outil à leurs besoins spécifiques.

[Le réseau des URFIST¹²](#) organise de nombreuses formations sur la gestion des données et les logiciels disponibles.

Des guides

[Les guides du FSD¹³](#) s'intéressent à tous les aspects de la gestion des données, y compris pratiques, par exemple le format de fichier ou les modalités pour le stockage. Ils sont disponibles en anglais ou en finnois.

[Le Vademécum pour la réutilisabilité des données](#) issu de l'activité du groupe de travail « (Ré) utilisabilité » au sein du consortium Cahier propose des bonnes pratiques et des astuces facilitant la gestion des données en vue d'une future réutilisation.

[Les guides du CIRAD¹⁴](#) présentent le PGD et centralisent des informations sur la science ouverte et son environnement.

11. Portail d'outils et de services facilitant la mise en application des principes FAIR.

12. Unité régionale de formation à l'information scientifique et technique.

13. Finish Social Sciences Data Archive.

14. Centre de coopération internationale en recherche agronomique pour le développement.

Contacts

L'équipe projet du PPR Autonomie se tient à votre disposition pour vous orienter vers des ressources ou des personnes compétentes. Des actions sont conduites afin de favoriser l'émergence et le développement d'une communauté de pratiques autour des expériences de gestion, diffusion et réutilisation des données.

N'hésitez pas à nous suivre pour rester informé de nos actualités !



[LinkedIn](#)



[Site du PPR Autonomie](#)



[S'inscrire à la newsletter du PPR Autonomie](#)



[Contacter l'équipe projet](#)

[Les guides de la bibliothèque de Sciences Po](#) offrent des outils pour rédiger un PGD ou aider à arbitrer, par exemple sur les outils de stockage. Ils proposent aussi un plan de PGD.

Des outils

[Nesstar Publisher](#) documente les données d'enquêtes par questionnaire dans un format DDI¹⁵.

[QAmYDATA](#) permet de vérifier si les fichiers stockés dans un dossier correspondent bien à un certain nombre de critères de forme.

[Heurist](#) est un système d'information qui permet de documenter ses données et jeux de données et d'établir des relations entre ces différents objets. Il possède aussi des fonctionnalités de visualisation du corpus, par exemple de géolocalisation.

[OpenRefine](#) pour mettre en forme des données quantitatives.

Plus de ressources

Au sein des [guidelines de la bibliothèque de Sciences Po](#).

Dans la [bibliographie et webographie de DORANum](#).

Sur le catalogue [CAT OPIDoR](#).

15. *Data Documentation Initiative*, format standard de documentation d'enquête.