

# Rédiger un Plan de Gestion des Données en 9 points

1. **Qu'est ce qu'un PGD**
2. **Pourquoi rédiger un PGD**
3. **Structure globale du PGD**
4. **Décrire les données**
5. **Clarifier le cadre éthique et juridique**
6. **Garantir la compréhension et l'accessibilité des données**
7. **Assurer le stockage, la sécurité et la sauvegarde des données**
8. **Définir les responsabilités**
9. **Spécifier les modalités de diffusion et de partage des données**

## 1. Qu'est ce qu'un PGD

Un PGD est un document qui explicite, dans un projet ou une thèse, la manière dont sont obtenues et traitées les données tout au long de leur cycle de vie, de la collecte à l'archivage. C'est un document formalisé, assez court.

L'enjeu n'est pas de décrire toutes les étapes mais de montrer que les données sont gérées selon des « **bonnes pratiques** » (de collecte, organisation, stockage, description, documentation, etc.), dans le respect d'un « cadre éthique et juridique », pour produire des données « Faciles à trouver, Accessibles, Interopérables and Réutilisables » (Principes FAIR : *Findable, Accessible, Interoperable and Reusable*), selon le principe « *aussi ouvert que possible aussi fermé que nécessaire* ».

Dans un projet, **un seul PGD** doit être rédigé pour donner une vue d'ensemble sur la gestion de toutes les données produites dans le projet. Le PGD doit donc résulter d'un **travail collectif** puisqu'il concerne tous les partenaires du projet qui collectent ou produisent des données. L'idéal est de nommer un responsable pour animer et coordonner la rédaction du PGD au sein du collectif. Le PGD implique une réflexion collective sur les dimensions juridiques et éthiques de la production et du partage des données et sur les conditions techniques de leur stockage, archivage et diffusion. Pour le porteur de projet et le collectif, le PGD représente, à la fois :

- une démarche de mise en œuvre de bonnes pratiques et de transparence des recherches
- un outil d'animation facilitant l'harmonisation des données et les échanges entre partenaires et assurant un gain de temps pour publier et valoriser les données et les résultats.

Le PGD décrit **les jeux de données** qui seront produits dans un projet. La définition du jeu de données est :

- un lot techniquement homogène (Ex : données de granulométrie, mesures de CO2, séquences)
- un lot intellectuellement cohérent même si composé de lots techniquement hétérogènes.

Il n'y a pas de règle imposée : c'est le collectif du projet qui détermine ce qui constitue un jeu de données homogène ou cohérent. Les jeux de données peuvent être bruts ou traités, disciplinaires (Ex : données économiques, de phénotypage, écologiques) ou techniques (Ex : données NIRS, données de séquençage, GPS, données issues de drones), par espèce (Ex : riz, sorgho, zébus), localisation (Ex : Cameroun, Vietnam, Nicaragua), ou work package, etc.

Le PGD est un **document évolutif**, il y aura donc plusieurs versions au cours d'un projet.

- La première version : dans les projets financés par **H2020** ou **l'ANR**, par exemple, la V1 doit être soumise au mois 6 du projet. Dans la V1, vous n'êtes pas obligé de répondre à toutes les questions. En effet, d'une part vous n'avez pas encore toutes les réponses et d'autre part la stratégie du projet en matière de conservation, publication et partage de certains jeux de données n'est pas complètement définie. Néanmoins, la V1 pose certaines questions pour vous aider à anticiper des problèmes d'ordre juridique, éthique ou technique dont la prise en compte et le traitement faciliteront la bonne gestion et la publication future des données.
- Les versions ultérieures (à mi-parcours pour l'ANR et H2020) seront complétées en fonction de l'évolution du projet et de sa stratégie pour les données produites.
- La version finale mentionnera les actions concrètes prévues par le consortium du projet pour respecter les principes FAIR et faciliter l'accès aux données et leur réutilisation éventuelle.

## 2. Pourquoi rédiger un PGD

Rédiger un PGD offre de **nombreux avantages** tels que :

- Garantir la qualité des recherches et la production de données fiables et compréhensibles
- Contribuer à la transparence, à l'intégrité scientifique et à la reproductibilité des recherches
- Assurer la sécurité des données personnelles, sensibles ou stratégiques
- Clarifier les rôles, responsabilités et droits de chaque contributeur
- Anticiper les problèmes juridiques, éthiques ou techniques
- Faciliter le partage de données au sein d'un collectif de projet
- Faciliter la publication d'articles de recherche dans des revues exigeant l'accès aux données
- Favoriser la mise à disposition de jeux de données à fort potentiel de réutilisation.

Rédiger un PGD répond aux demandes des tutelles, des institutions, des financeurs de projets, et aussi de certains éditeurs. Lorsqu'ils proposent un modèle de PGD, il est recommandé de le suivre.

## 3. Structure globale du PGD

La plupart des modèles de PGD suivent la même structure en **6 sections** : Description des données ; Documentation et métadonnées ; Éthique et droits de propriété intellectuelle ; Responsabilités ; Stockage et conservation ; Partage et réutilisation. Chaque section contient plusieurs questions (voir ci-dessous). Elles sont distribuées différemment selon les modèles mais ce sont souvent les mêmes.

En pratique, vous devez rédiger un seul PGD pour montrer la cohérence et le lien entre tous les types de données produits dans le projet. Pour cela, vous avez le choix de traiter soit :

- l'ensemble des types de données dans chaque section du PGD
- chaque type de données séparément en déclinant les 6 sections du PGD pour chacun d'eux.

Il n'y a pas de règle imposée : c'est un choix à faire en consultation avec les partenaires pour faciliter la rédaction à laquelle ils devront participer.

## 4. Décrire les données

**Dans le PGD, la première étape consiste à :**

- identifier et lister les types de jeux de données qui seront collectés ou générés dans le projet (Ex : données expérimentales, données d'observation, données d'enquêtes, de santé, de génomique, de phénotypage, écologiques, spatiales, de simulation, logiciels, images, etc.).
- spécifier l'objet d'étude, l'origine géographique, la période temporelle, etc.
- expliquer l'objectif de la collecte/génération de ces différents types de données.

## 5. Clarifier le cadre éthique et juridique

La 2eme étape consiste à clarifier le cadre éthique et juridique qui s'applique aux données. Celui-ci varie selon les données collectées, produites ou utilisées.

**Dans le PGD, il faut expliquer les mesures prises dans le cas où vos données contiennent :**

- des **données personnelles** (Ex : données récoltées lors d'enquêtes ou de questionnaires)  
*Dans ce cas, expliquez les mesures appliquées pour satisfaire au [règlement général sur la protection des données \(RGPD\)](#) qui impose des mesures de confidentialité, le consentement éclairé des participants, la sécurisation des données, l'anonymisation des données, etc.*  
Pour plus d'informations : [https://www.inshs.cnrs.fr/sites/institut\\_inshs/files/pdf/guide-rgpd\\_2.pdf](https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf)
- des données qui soulèvent des **questions éthiques** (Ex : lors d'expérimentation animale, de suivi de traitements dans des populations, de recherches pouvant avoir un impact sur l'environnement, la conservation de la biodiversité, la santé des chercheurs impliqués, etc.).  
*Dans ce cas, expliquez comment vous traitez ces questions éthiques, si vous avez sollicité l'approbation d'un comité d'éthique approprié, etc.*  
Outil d'aide pour les questions éthiques : <https://theodi.org/article/data-ethics-canvas/>
- des données issues de **ressources génétiques** ou de savoirs traditionnels associés.  
*Dans ce cas, et selon le pays fournisseur des ressources génétiques, précisez les procédures mises en place pour respecter la [législation APA \(Accès et Partage des Avantages\)](#).*  
Pour connaître la législation du pays fournisseur : <https://absch.cbd.int/fr/countries>.
- des **données préexistantes**, c'est-à-dire produites avant ce projet, par vous (Ex : dans le projet X qui est terminé), par d'autres (Ex : collègues de l'université Y) ou qui sont accessibles dans des entrepôts ou des observatoires (Ex : FAOSTAT, GBIF, Centre for Ecology & Hydrology).  
*Dans ce cas, mentionnez la source à l'origine de ces données préexistantes et précisez si elles sont libres d'utilisation ou protégées par des droits spécifiques ou des licences. Si aucune donnée n'est disponible, précisez-le car cela justifie le besoin de produire ces données pour répondre à votre question de recherche.*
- des données pour lesquelles des **droits de propriété intellectuelle** sont à considérer (Ex : images protégées par droit d'auteur ou données obtenues en partenariat ou sous contrat avec un acteur privé ou utilisation de données préexistantes sous licence, ...) ?  
*Dans ce cas, détaillez les jeux de données concernés et les droits de propriété intellectuelle à considérer et précisez ce qui est décrit dans l'accord de consortium du projet.*

## 6. Garantir la compréhension et l'accessibilité des données

La 3eme étape consiste à expliquer comment les données seront documentées, décrites, organisées et formatées. L'objectif est d'assurer la compréhension des données, de faciliter la reproductibilité des recherches et de permettre l'accès et la réutilisation des données. Vous devez utiliser, autant que possible, des méthodes, formats, unités et standards de description qui sont classiques de la discipline et seront donc compris par la communauté scientifique concernée (et si possible au-delà).

Pour trouver les normes et standards disciplinaires qui existent :

Data Curation Center : <http://www.dcc.ac.uk/resources/metadata-standards>

Research Data Alliance : <http://rd-alliance.github.io/metadata-directory/standards/>

FAIR Sharing : <https://fairsharing.org/>

Elixir : <https://elixir-europe.org/communities>

Pour être sûr de bien documenter vos données, mettez-vous à la place d'un futur utilisateur et demandez-vous si vous avez fourni toutes les informations indispensables pour que vos données soient accessibles, lisibles, compréhensibles et réutilisables.

**Dans le PGD, vous devrez :**

- décrire les méthodes qui sont utilisées (Ex : protocoles d'échantillonnage, de collecte, de production des données, protocole d'enquête, types d'équipement, explication des variables étudiées, etc.) et préciser s'il s'agit de méthodes classiques dans la discipline.
- mentionner les processus de contrôle qualité pour montrer la rigueur des méthodes et la qualité des données (Ex : processus de calibration, mesures répétées, contrôles standards positifs/négatifs, contrôle des données en double aveugle ou par évaluateur externe, etc.).  
préciser tous les documents qui seront associés aux fichiers de données (Ex : questionnaires d'enquête, protocoles, dictionnaires des variables, abréviations, codes, versions des logiciels de lecture, fichier « Lisez-moi », ...). L'objectif est que chaque fichier de données soit accompagné de toutes les informations nécessaires pour que les données puissent être comprises et interprétées par quelqu'un qui n'a pas participé à l'étude. Préciser également si des articles associés à ces données ont déjà été publiés et si oui, donner les références.
- montrer que vous utilisez les normes et standards de description des données (métadonnées) pratiqués dans la discipline, lorsque ceux-ci existent (Ex : norme ISO 8601 pour les dates (AAAA/MM/JJ) ; norme ISO 19115 pour une géolocalisation ; standard EML pour des données en écologie ; norme DDI pour des données d'enquêtes, standard MIAPPE pour du phénotypage de plantes ; MlXs pour les séquences ADN ; etc.). Dans le cas où ces standards n'existent pas dans votre discipline, expliquez quels éléments descriptifs (métadonnées) vous utilisez et comment ils sont produits (Ex : cahier de laboratoire, GPS, autre type d'instrument, entrée manuelle) ?
- préciser l'organisation, les règles de nommage et le format des fichiers. Privilégier les formats usuels de la discipline (Ex : Excel, csv, FASTA, BAM, JPEG, SPSS, Gis, etc.) et surtout les formats « ouverts » pour faciliter l'accès et la réutilisation des données. Si des logiciels ou outils sont nécessaires pour lire les données, précisez-le (Ex : Excel, R, Matlab, Nesstar, etc.).

## 7. Assurer le stockage, la sécurité et la sauvegarde des données

La 4ème étape consiste à décrire pour chaque type de données les supports de stockage et les procédures de sauvegarde et sécurisation (s'il s'agit de données personnelles ou sensibles) appliquées pendant et après le projet pour éviter tous risques de brèches et perte de données (Ex : ordinateurs portables, protection par mot de passe, disque dur externe, clés USB, plateformes collaboratives, serveur local, serveur distant, entrepôts de données, cloud, etc.).

Vous devrez aussi réfléchir, compte tenu des coûts que représente l'archivage de données, aux jeux de données qui seront archivés parce qu'unique et à ceux qui seront détruits à la fin du projet parce que facilement reproductibles.

**Dans le PGD, vous devrez :**

- préciser la localisation de ces supports et la fréquence de sauvegarde. La règle basique est celle du 3-2-1 : 3 copies des données (et documents) sur 2 supports différents (Ex : clé USB, disques durs externes, serveur institutionnel, etc.) dont 1 stocké dans une localisation différente (Ex : hors de son bureau).
- expliquer le mode de protection des données si vous gérez des données personnelles, sensibles, ou stratégiques et les procédures de contrôle des droits d'accès (voir : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre4#Section2>)
- évaluer le volume de données envisagé (Ex : 50 Mo, 30 Go, 1 To) et préciser si l'espace de stockage dont vous disposez est suffisant.

- évaluer les coûts financiers (humain et technique) pour la gestion et le stockage des données du projet (voir: <https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>). Pour information, les coûts de gestion des données sont éligibles chez de nombreux financeurs.
- spécifier les données qui méritent d'être conservées sur le long terme et expliquer les critères de sélection, et celles qui pourront être détruites après le projet pour des raisons de coûts ou d'exigences réglementaires (Ex : données à caractère personnel conformément aux recommandations de la CNIL et du RGPD)?

## 8. Définir les responsabilités

Cette partie consiste à nommer toutes les personnes qui auront une responsabilité dans la gestion des données, pour : la collecte, le traitement, l'analyse, le stockage, l'anonymisation, la rédaction des versions du PGD, etc. Selon le projet, les responsabilités peuvent également être définies par localisation, par work-package, par espèce traitée, etc.

**Dans le PGD, vous devrez :**

- mentionner les nom(s), UR, institution et e-mail(s) de toutes les personnes qui portent une responsabilité dans une activité liée à la gestion des données.

## 9. Spécifier les modalités de diffusion et de partage des données

La dernière étape consiste à décrire 1) les modalités d'échanges de données pendant le projet, notamment entre partenaires et 2) les modalités d'accès, de diffusion et de partage des données, pendant le projet puis après le projet. Ces modalités peuvent varier selon les types de données, le volume, l'origine des données, les pratiques des disciplines concernées, etc.

Le point 1 doit être traité dans la V1 du PGD car il s'agit de procédures qui doivent être anticipées en début de projet. Ces procédures représentent les bonnes pratiques qui sont attendues dans le PGD.

**Dans le PGD, vous devrez :**

- préciser si des échanges de données sont prévus entre partenaires pendant le projet. Si oui, avec quels partenaires (internes, externes) ? comment ? (Ex : plateforme collaborative institutionnelle, plateforme open source), selon quelles modalités d'accès ? (Ex : accès pour tous les partenaires ou restreint à certains) et quelle procédure, notamment en terme de transfert sécurisé de données s'il y a des données personnelles ou sensibles ?
- préciser si des échanges sont également prévus avec des personnes extérieures au projet.

Le point 2 est plus difficile à traiter en début de projet. Il est vraisemblable que vous n'avez pas encore toutes les réponses concernant la diffusion des données produites par le projet. Rappelez-vous qu'il y aura plusieurs versions du PGD et que vous n'êtes pas obligé de répondre à toutes les questions dans la V1. Précisez néanmoins que ces questions ont été prises en compte par les partenaires du projet et que la stratégie de gestion et de partage des données sera affinée au cours du projet. Sachez qu'à minima, vous êtes censés partager les données qui étayent les articles de recherche publiés par le projet.

Pour aborder ce point 2, posez-vous les questions suivantes :

- Etes-vous soumis à une obligation de partage des données par votre institution, votre partenariat, le financeur du projet, l'éditeur chez qui vous souhaitez publier, etc. ?
- Etes-vous soumis à une restriction de partage des données par l'un de vos partenaires, le financeur du projet, la source utilisée pour récupérer des données préexistantes, etc. ?
- Certains jeux de données contiennent-ils des données personnelles, confidentielles, sensibles ou contractuelles qui empêchent de les partager ? avez-vous signé un accord de confidentialité ? avez-vous obtenu le consentement libre et éclairé des personnes dont vous utilisez les données ? les données sont-elles anonymisées ?

- Certains jeux de données sont-ils uniques ? (Ex : expérimentations impossibles à répéter, groupes difficilement accessibles, phénomènes rares), méritent-ils d'être valorisés en les rendant accessibles et réutilisables dans un entrepôt de données ?
- Certains jeux de données présentent-ils un fort potentiel de réutilisation et un intérêt pour d'autres scientifiques ? pour la société civile ? pour des décideurs ? pour le secteur privé ? Quelles réutilisations possibles ? Quelles perspectives d'application ?

**Dans le PGD, vous devrez :**

- dire si vous avez l'intention de partager tous les jeux de données produits par le projet ou seulement certains jeux de données
  - spécifier les jeux de données qui seront partagés (*a minima*, ceux qui étayent les articles publiés par le projet)
  - expliquer les raisons qui justifient de ne pas partager les autres jeux de données (Ex : présence de données confidentielles, sensibles, stratégiques ou contractuelles).
  - préciser quand, où et comment seront partagés les jeux de données. Ces informations peuvent être différentes selon les jeux de données :
    - quand ? (Ex : en année 3 du projet pour les données génomiques, à la fin du projet pour les données biochimiques, juste après publication de l'article pour les données sur le Cameroun, après un embargo de 2 ans pour les données de phénotypage du riz pour pouvoir publier 2 articles, etc.)
    - où ? La procédure recommandée est le dépôt dans un entrepôt de données institutionnel, thématique ou disciplinaire (Ex : Dataverse, GenBank, PANGAEA, etc.). Répertoire d'entrepôts de données : <http://www.re3data.org/>
    - selon quelles modalités d'accès ? (Ex : accès libre, accès après enregistrement, accès restreint, accès uniquement sur demande, accès différé après embargo de 2 ans)
    - avec quelle licence de diffusion ? (Ex : licence *Creative Commons* de type CC-BY, licence avec exclusion d'utilisation commerciale (CC-BY-NC), Open database License).
- Pour plus d'informations : CoopIST : Rendre publics ses jeux de données – Les principales licences de diffusion <https://doi.org/10.18167/coopist/0059>

## Liens utiles

### Politiques des agences de financement concernant le PGD

Guide pour les projets H2020 : [Guidelines on FAIR Data Management in H2020](#)

Guide pour les projets ANR : [https://www.ouvrirlascience.fr/wp-content/uploads/2019/07/Plan-de-gestion-de-donn%C3%A9es-Recommandations-%C3%A0-ANR\\_vdef.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2019/07/Plan-de-gestion-de-donn%C3%A9es-Recommandations-%C3%A0-ANR_vdef.pdf)

Guide pour les projets B & M Gates Fondation : <https://gatesopenresearch.org/documents/2-46>

Sherpa-Juliet, pour connaître la politique des financeurs : <http://v2.sherpa.ac.uk/juliet/>

### Modèles de PGD

Outil de rédaction de PGD : <https://dmp.opidor.fr/>

Data Management Plan (DMP). Ecole Polytechnique Fédérale de Lausanne (EPFL). <https://researchdata.epfl.ch/plan-fund/dmp/>. Voir le modèle (*template*) disponible à partir de cette page, il est en anglais mais très complet et il contient de nombreux exemples.

N. Reymonet, M. Moysan, A. Cartier, R. Délémontez. Réaliser un plan de gestion de données "FAIR" : modèle. 2018. sic\_01690547v2. [https://archivesic.ccsd.cnrs.fr/sic\\_01690547/document](https://archivesic.ccsd.cnrs.fr/sic_01690547/document)

### Cadre juridique des données de recherche

Ouverture des données de recherche. Guide d'analyse du cadre juridique en France. V2. 2017. [http://corist-shs.cnrs.fr/sites/default/files/ressources/guide\\_analyse\\_cadre\\_juridique\\_ouverture\\_donnees\\_recherche\\_v2\\_lienceouverte\\_prefacedgri.pdf](http://corist-shs.cnrs.fr/sites/default/files/ressources/guide_analyse_cadre_juridique_ouverture_donnees_recherche_v2_lienceouverte_prefacedgri.pdf)

Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. Guide pour la recherche. InSHS-CNRS. 2019.

[https://www.inshs.cnrs.fr/sites/institut\\_inshs/files/pdf/guide-rgpd\\_2.pdf](https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf)

Réglementation européenne sur l'accès et le partage des avantages (APA). Fondation pour la Recherche sur la Biodiversité. 2019. <https://www.fondationbiodiversite.fr/reglementation-europeenne-sur-laces-et-le-partage-des-avantages-apa/>

Outil d'aide à la décision. INRA. <http://www.bibliotheque-numerique.fr/DonneesDiffusables.php>

### Exemples de PGD accessibles sur internet

DMP Online : accès à tous les PGD publics : [https://dmponline.dcc.ac.uk/public\\_plans](https://dmponline.dcc.ac.uk/public_plans)

DMP Tool: [https://dmptool.org/public\\_plans?e=f&public%3Apage=1&s=a](https://dmptool.org/public_plans?e=f&public%3Apage=1&s=a)

PGD publiés dans la revue RIO (Editeur Pensoft)

[https://riojournal.com/browse\\_journal\\_articles.php?form\\_name=filter\\_articles&sortby=0&journal\\_id=17&search\\_in\\_=0&section\\_type%5B%5D=231](https://riojournal.com/browse_journal_articles.php?form_name=filter_articles&sortby=0&journal_id=17&search_in_=0&section_type%5B%5D=231)

Comment citer ce document :

Dedieu, L. 2019. Rédiger un Plan de Gestion des Données en pratique. Montpellier (FRA) : CIRAD, 7 p.

<https://doi.org/10.18167/coopist/0066>

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International, disponible en ligne : <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.fr> par courrier postal à : Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA. Cette licence vous permet de remixer, arranger, et adapter cette œuvre à des fins non commerciales tant que vous créditez l'auteur en citant son nom et que les nouvelles œuvres sont diffusées selon les mêmes conditions.