

## Le Plan de gestion de données (PGD)

### 1- Définition :

Le **PGD - Plan de gestion de données** est appelé aussi **DMP pour Data Management Plan**.

Le PGD est un **outil de gestion**. Il se présente sous la forme d'un document structuré en rubriques et a pour objectif de synthétiser la description et l'évolution des jeux de données de votre projet de recherche.

Il prépare ainsi le partage, la réutilisation et la pérennisation des données.

Le PGD est devenu un **document incontournable** :

- **À l'échelle mondiale :**

Le PGD est devenu un outil de pilotage très répandu. Il est de plus en plus recommandé ou exigé, partout dans le monde.

- **À l'échelle européenne :**

La rédaction d'un PGD est une **exigence de la Commission européenne** : la version initiale du PGD est inscrite parmi les livrables à 6 mois après le début du projet (Modèles Horizon 2020, Horizon Europe, ERC-European Research Council).

Pour favoriser la gestion et le partage des données de la recherche, de nombreuses **initiatives d'ampleur européenne** ont été déployées, notamment des **outils** et des **infrastructures**, par exemple :

- l'entrepôt [Zenodo](#),
- l'infrastructure [OpenAIRE](#) (Open Access Infrastructure for Research in Europe),
- l'initiative [EOSC](#) (European Open Science Cloud)...

- **À l'échelle nationale :**

Le financeur ANR (Agence nationale de la recherche) a rendu obligatoire la rédaction d'un PGD depuis 2019. Il propose un modèle de PGD ANR en version française et anglaise.

L'**État français** a élaboré une **politique nationale** avec un premier [Plan national pour la science ouverte](#) (juillet 2018).

Après un bilan positif, cette politique nationale a été précisée et renforcée dans un [Deuxième Plan national pour la science ouverte](#) (juillet 2021).



Le PGD est aussi inscrit à l'**article 6** du **décret n°2021-1572** du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique.

<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>

- **À l'échelle des organismes :**

De nombreux instituts, organismes, établissements mettent à disposition de leur communauté des modèles de PGD institutionnels (modèles de PGD du CIRAD, de l'INRAE, de l'Institut Pasteur, de l'université de Strasbourg, de Sciences Po...).

Certains établissements comme l'INRAE ont mis en place une politique d'établissement pour la gestion et le partage des données de recherche.

De plus en plus d'établissements émettent des recommandations institutionnelles.

- **À l'échelle disciplinaire :**

Des modèles de PGD disciplinaires commencent à être mis en place.

C'est le cas en astronomie, en archéologie (modèle du consortium MASA, disponible dans l'outil DMP OPIDoR)...

## **2- Acteurs et contributeurs :**

Le PGD est une opportunité de dialogue entre les différents acteurs d'un projet : scientifiques, informaticiens, professionnels de l'IST (Information Scientifique et Technique), chargés de projet, juristes... sans oublier les partenaires du projet.

Le PGD est en effet l'occasion de fixer dès le départ les règles de gestion des données entre les différents partenaires d'un projet et de cadrer les échanges avec des partenaires privés. Il peut s'associer aux accords de consortium notamment.

Le chercheur n'est donc pas seul face à la rédaction du PGD. La gestion des données demande un **effort collectif** !

- **La communauté scientifique :**

Les chercheurs du même domaine peuvent être d'une grande aide grâce au partage d'expérience, par exemple dans le choix de standards de métadonnées ou celui d'un entrepôt de données...

- **L'ingénieur-projet :**

Il aide le chercheur à coordonner les actions autour du PGD mais aussi dans l'obtention d'un agrément, pour l'éligibilité des coûts...

- **Service éthique et juridique :**

Le chercheur peut solliciter le service juridique, notamment pour être conseillé sur la propriété intellectuelle des données.



Dans les universités, il est aussi souvent possible de se faire accompagner sur ces questions par un comité éthique.

- **Délégué à la protection des données :**

Le délégué à la protection des données est l'interlocuteur idéal pour aider le chercheur sur les aspects RGPD (règlement général sur la protection des données).

- **Professionnel de l'IST :**

Les professionnels de l'Information Scientifique et Technique (IST), notamment les documentalistes, peuvent aider le chercheur dans le choix des standards de métadonnées, conseiller sur les entrepôts mais aussi réaliser des alignements avec des référentiels existants...

- **Informaticien :**

C'est l'interlocuteur privilégié en ce qui concerne le stockage et la sécurisation des données, les aspects infrastructure et les coûts associés.

- **Archiviste :**

Il peut aider le chercheur à sélectionner les données pour la conservation, à définir les durées et les solutions techniques.

- **Chercheur :**

Il est la personne qui connaît le mieux les données et le projet. Il est donc généralement le coordinateur du PGD et le responsable des données. Il lui revient de décrire les données, de procéder au découpage des jeux de données...

Les universités, infrastructures et organismes de recherche émettent souvent des **recommandations** à destination de leurs communautés de recherche. De même les financeurs (comme l'ANR ou la Commission européenne) et certains éditeurs peuvent donner des **consignes** précises (par exemple : obligation de rédaction d'un PGD dans les 6 mois suivant le début du projet pour les projets financés par l'ANR) ou proposer des **conseils** (par exemple : la Commission européenne indique l'existence de l'entrepôt Zenodo dans son guide Horizon 2020).

Le répertoire des **Services Opérationnels de Soutien à la rédaction des Plans de Gestion des Données (SOS-PGD)** recense les services accompagnant la rédaction des plans de gestion des données au sein des établissements d'enseignement supérieur et de la recherche.

Il vise à aider les chercheurs à identifier leurs interlocuteurs au sein de leur institution et à faciliter la mise en relation entre les services supports de différentes institutions pour les projets de recherche multi-partenariaux.

Voir le site : <https://scienceouverte.couperin.org/sos-pgd/>



### 3- Un outil de gestion de projet :

Le PGD est un document **évolutif**, dynamique et continuellement mis à jour (introduction d'un nouveau jeu de données, données faisant l'objet d'un dépôt de brevet, changement dans le consortium...).

C'est aussi un **outil de gestion de projet** qui aide à organiser ses données, à les décrire, à bien définir les responsabilités, les ressources nécessaires et à produire des données fiables.

- **Organisation des données :**  
Le PGD aide à bien organiser les données, tout au long du projet.
- **Document évolutif :**  
Il faut commencer à rédiger le PGD dès le début du projet, avec les éléments déjà connus ou prévus. Ensuite, compléter le PGD au fur et à mesure.  
Prévoir 2 versions au minimum : au début et à la fin du projet.  
Pour les projets de plus de 30 mois, une version intermédiaire est demandée.
- **Description des données :**  
Dans le PGD, il faut décrire la façon dont les données seront obtenues, traitées, organisées, stockées, sécurisées, partagées... (cycle de vie des données).
- **Données fiables :**  
Le PGD permet aux producteurs de données de se poser les bonnes questions et donc d'améliorer la fiabilité de leurs données.
- **Responsabilités :**  
Dans le PGD, désigner la ou les personne(s) responsable(s) de la gestion des données pour toutes les étapes du projet et au sein du partenariat s'il y a lieu :
  - saisie des données
  - production des métadonnées
  - contrôle de la qualité des données
  - stockage, partage et archivage des données
  - mise à jour du PGD.

On peut désigner des personnes nominativement ou indiquer une fonction si la personne qui l'occupe peut être amenée à changer au cours du projet.



- **Ressources :**

Il est demandé d'évaluer les ressources nécessaires (budget, temps alloué, personnels) permettant la mise en œuvre des actions décrites dans le PGD :

- temps nécessaire à la préparation des données pour le stockage, le partage et l'archivage des données
- coûts de matériel et rémunération des personnels
- frais de stockage (serveurs dédiés, traitement, maintenance, sécurité, accès...), frais de partage (site web, publication...) et frais d'archivage des données.

La rédaction d'un plan de gestion permet ainsi d'initier très tôt une **réflexion collective** sur les bonnes pratiques et d'**anticiper** les questions relatives à la gestion des données (comme le choix de l'entrepôt, la documentation à associer...).

#### **4- Différents modèles :**

Il n'existe pas de trame unique de PGD. Toutefois, de nombreux modèles ont été établis par des organismes, instituts, financeurs à destination de leurs utilisateurs. Ce peut être pour répondre aux spécificités propres à certains organismes de recherche, pour correspondre au contexte local des établissements, etc.

On retrouve néanmoins les mêmes éléments dans tous les modèles de PGD, à savoir :

- Informations administratives
- Description des données
- Documentation, métadonnées, standards
- Aspects juridiques
- Responsabilités
- Sécurité des données
- Stockage des données durant le projet
- Partage des données avec le dépôt dans un entrepôt
- Archivage pérenne
- Coûts

#### **Exemple : le modèle ANR.**

Ce modèle est composé de **6 grandes thématiques** illustrant les bonnes pratiques de gestion et de partage. Pour chaque thématique, le chercheur est invité à répondre à plusieurs questions.

##### **1. Description des données et collecte ou réutilisation des données existantes**

- a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?
- b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?



## 2. Documentation et qualité des données

- a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?
- b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

## 3. Stockage et sauvegarde pendant le processus de recherche

- a. Comment les données et métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?
- b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

## 4. Exigences légales et éthiques, codes de conduite

- a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?
- b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?
- c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

## 5. Partage des données et conservation à long terme

- a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?
- b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?
- c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?
- d. Comment l'application d'un identifiant unique et pérenne (comme le DOI) sera réalisée pour chaque jeu de données ?

## 6. Responsabilités et ressources en matière de gestion des données

- a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données ?
- b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données soient FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

## 5- DMP OPIDoR :

DMP OPIDoR est un **outil de rédaction** de PGD en ligne, hébergé en France et accessible gratuitement.

Il suffit de se créer un compte pour rédiger des PGD, **en anglais ou en français**, à partir du modèle choisi (nombreux modèles financeurs et établissements).

C'est un **outil collaboratif** qui permet de partager un PGD avec un ou plusieurs collaborateurs, français et étrangers : ceux-ci peuvent le **co-rédiger et/ou ajouter des commentaires** en fonction des droits qui leur sont accordés.



- **Recommandations :**

Chaque modèle de PGD est accompagné d'un ensemble de recommandations.

L'outil DMP OPIDoR offre la possibilité d'afficher les recommandations associées à plusieurs modèles de PGD à la fois. Pour cela, ouvrir l'onglet "Informations générales" : un encadré "Sélection des recommandations du plan" se trouve sur la partie droite de l'écran. Cliquer sur "Voir la liste complète" et cocher les organismes dont les recommandations vous intéressent.

- **Assistance conseil :**

Il est possible de demander une assistance conseil auprès des services d'appui de votre organisme de recherche. Attention, cette option n'est disponible que si votre organisme est doté d'un service d'accompagnement et si celui-ci a activé ce service.

**Attention :** DMP OPIDoR est uniquement un outil de rédaction. Il n'a pas été conçu pour servir d'entrepôt de PGD et n'a donc pas vocation à conserver les PGD sur le long terme.

Voir le site : <https://dmp.opidor.fr/>

