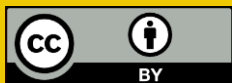


# Comment rédiger un Plan de gestion des données ?

Rédiger un PGD- 16 janvier 2023

SCD – Département d'Appui à la Recherche 2022-2023



Cycle de vie de la donnée et principes FAIR	1
Le contexte national et européen	2
Qu'est-ce qu'un plan de gestion des données ?	3
Description des données	4
Documentation et qualité des données	5
Stockage et sauvegarde	6
Exigences légales et éthiques	7
Partage des données et conservation à long terme	8
Responsabilités et ressources	9
Quelles aides pour la rédaction ?	10

# Plan



## Quelques notions importantes autour des données de la recherche

*Cycle de vie de la donnée*  
*Principes FAIR*

**Pour aller plus loin :**

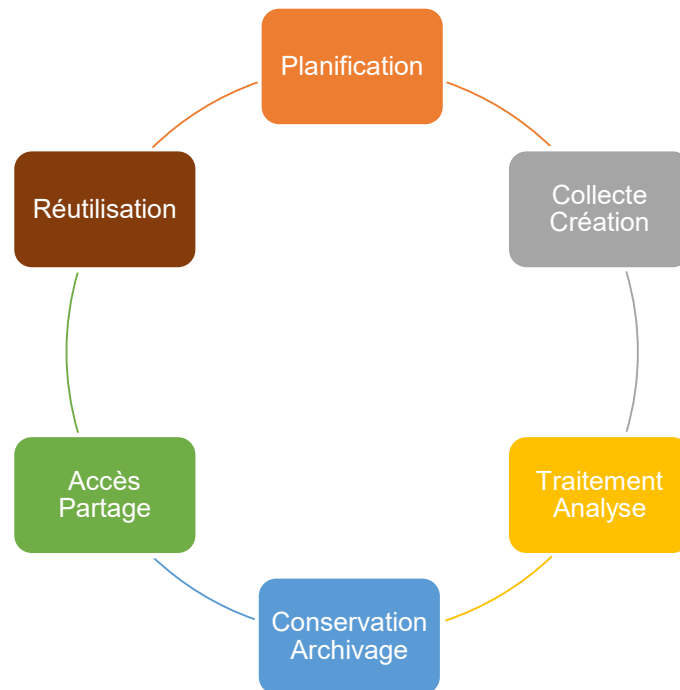
[Sur le cycle de vie de la donnée](#)

[Sur les principes FAIR](#)

## 1. Cycle de vie de la donnée

### Définition :

Ensemble des opérations réalisées tout au long du cycle de vie de la donnée pour en assurer la qualité et la pérennité.



# 1. Les principes FAIR

Des principes qui garantissent une bonne gestion des données de la recherche...

## Findable



Facile à trouver

- Descriptions
- Identifiant pérenne

## Accessible



Facile à utiliser

- Protocole de communication ouvert
- Authentification
- Accès aux métadonnées

## Interoperable



Facile à échanger

- Standards de description
- Format ouvert
- Données liées

## Reusable



Facile à réutiliser

- Licence de réutilisation
- Provenance
- Format ouvert/standard

## 2. Le contexte national et européen

### Le contexte national et européen

#### *Cadre européen*

Agences européennes encouragent fortement le dépôt des données issues des travaux qu'elles financent.

#### **Quelles obligations ?**

La rédaction d'un plan de gestion des données.

L'accès libre aux données et aux publications en respectant le principe :

*« Aussi ouvert que possible, aussi fermé que nécessaire »*

Les données doivent être déposées dans un entrepôt de confiance.



## 2. Le contexte national et européen

### Le contexte national et européen

*Cadre national*

#### Politique Science Ouverte de l'ANR :

Impose la rédaction d'un PGD pour tout projet de recherche financé (2019).

L'ouverture des données n'est pas obligatoire, MAIS encouragée et souhaitée.

**Second plan national science ouverte (2021-2024)** : structuration, partage et ouverture des données de la recherche.

**Juillet 2022** Lancement de l'entrepôt national des données de la recherche, Recherche Data Gov.



RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*

recherche.data.gouv.fr

### 3. Qu'est-ce qu'un plan de gestion des données ?

#### Le Plan de gestion des données

*Data Management Plan*

Outil essentiel dès le début des recherches, il expose comment seront gérées les données pendant et après le projet.

S'intéresse à l'ensemble du cycle de vie des données et est mis à jour et enrichi au fur et à mesure.

Sert à formaliser et partager ses pratiques sur la gestion des données durant l'ensemble de leur cycle de vie.

Document partagé avec toutes les parties prenantes du projet de recherche.

Livrable attendu par l'ANR et la commission européenne **6 mois après le début du projet.**





### 3. Qu'est-ce qu'un plan de gestion des données ?

Il existe aussi des :

**PGL\*** : Plan de gestion de logiciel.

*\*Pour réfléchir et planifier le développement et la gestion de code.*

1. Métadonnées
2. Contexte du logiciel
3. Caractéristiques du logiciel
4. Organisation de l'équipe
5. Organisation du développement
6. Organisation de la diffusion
7. Gestion du plan de gestion

[Modèle de PGL](#)



## 3. Qu'est ce qu'un plan de gestion des données ?

### Le Plan de gestion des données

*Data Management Plan*

Le modèle de PGD de l'ANR

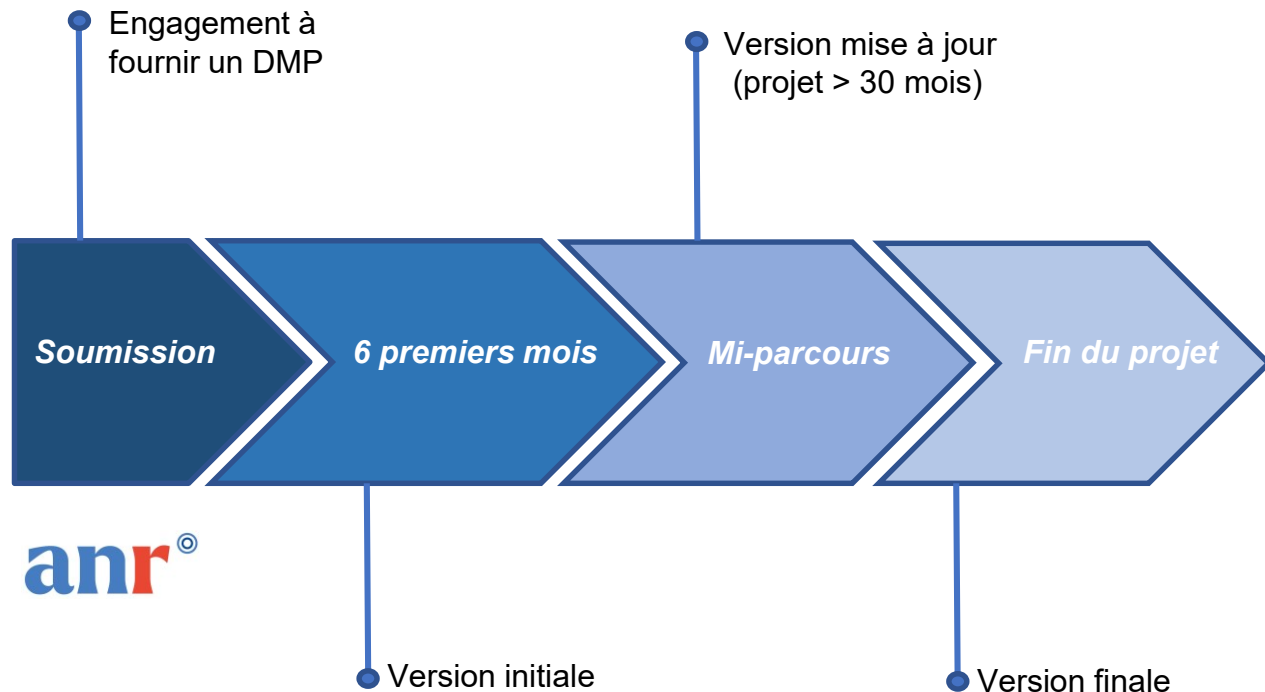
1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données et conservation à long terme
6. Responsabilités et ressources

Le modèle de PGD Science Europe

1. Data Summary.
2. FAIR data
3. Other research outputs
4. Allocation of ressources
5. Data Security
6. Ethics
7. Other issues

# Présentation du modèle ANR

## 1 DMP par projet.



# Présentation du modèle ANR



# 1.Description des données

## Pourquoi décrire ses données ?



Cette première partie est indispensable pour rédiger l'ensemble du PGD.

La description des données est là pour vous aider à identifier les principaux enjeux auxquels vous allez être confrontés :

- Le stockage
- L'accès
- La nature des données
- La pérennité...

En décrivant de manière précise vos données, vous les rendrez compréhensibles et utilisables par les membres du projet, voire d'autres chercheurs.



# 1. Description des données

On peut s'interroger sur les questions suivantes :

- Quelles sont les données produites, collectées, réutilisées ?
- Quelles sont leurs caractéristiques ?

**Types** : Vidéo, audio, textes, codes...

**Source / Provenance** : Collecte de nouvelles données, conversion ou transformation des données, partage/échange de données, achat de données etc.)

**Formats** : Privilégier des formats ouverts et non propriétaires (Jpeg, odt, tiff, pdf).

**Volume** : Quantité de données, octets, terra octets.

**Stabilité** : Jeu de données achevé, croissant.

**Valeur** : Coût financier, en temps, en moyens.

**Propriété** : Données réutilisées ou non, originales.

**Sensibilité** : Données confidentielles, sensibles, médicales.

## Conseils de rédaction

Un jeu de données à décrire = un « lot » de données pour un type de données défini.

Réaliser des **sous-ensembles** de données pertinents selon...

- Le type de données
- La provenance
- La propriété
- Outils/Méthodes employés

Exemple provenant du DMP public, *Hospitam Lauriane Locatelli (ENS Lyon)* :

« Les données produites seront les suivantes :

- Documents **textuels** des textes antiques non lemmatisés et lemmatisés en langue grec (**alphabet grec ancien avec accent, format Unicode**) en **.txt** regroupé en dossiers par auteurs, c'est-à-dire **6 dossiers de maximum 30 Mo** [...] »

## Conseils de rédaction

Pour chaque partie du PGD décliner les réponses en **reproduisant le classement**.

Par exemple :

*DataSet1 Images de microscopie de cellules apoptotiques*

*Type : Images*

*Outil: Microscope*

**Quelques ressources...**

**[Description des données](#)**

**[Liste de formats ouverts et fermés](#)**



## 2.Documentation et qualité des données

**La documentation et les métadonnées** permettent de garantir que les données soient **compréhensibles et réutilisables**.

Les métadonnées sont des données sur les données :

- Permettent de mieux organiser et gérer ses données
- Facilitent la compréhension des données
- Favorisent la pérennisation

***Pour aller plus loin sur cette question :***

[Metada Standards Catalog](#)

[Des schémas de métadonnées](#) (vidéo)

[Standards de métadonnées](#)

[Règles de nommage de documents](#)

## 2.Documentation et qualité des données

### On s'interroge sur :

1.Quelles sont les conditions de **production** ou de **collecte** des données ?

- Méthodologie, protocoles expérimentaux, traitements appliqués, logiciels ou instruments utilisés...

2.Comment les données du projet sont-elles **décrites** ? Des **standards de métadonnées** sont-ils utilisés ?

- Dublin core, XML, référentiel disciplinaire
- [Metada Standards Catalog](#), un outil pour trouver un standard adapté à sa discipline.

3.Où les métadonnées se trouvent-elles ?

- Métadonnées embarquées (exemple : image).
- Métadonnées externes : fichier « readme », dictionnaire des codes, cahier de laboratoire...

## 2.Documentation et qualité des données

- Comment les données sont-elles **organisées** ?  
*Organisation des dossiers*  
*Règles de nommage des fichiers*
- Comment la **qualité** des données est-elle assurée ?  
*Contrôle de la cohérence des données, taille ou répétition des échantillons, calibrage des instruments...*

### Conseils de rédaction :

- Produire de la documentation au fur et à mesure du projet.
- Amorcer une réflexion dès le début avec les membres du projet pour se concerter sur la documentation, le nommage des fichiers, l'organisation des dossiers partagés etc...
- Se renseigner sur les pratiques disciplinaires en matière de schémas de métadonnées.

## 2.Documentation et qualité des données

### Focus Fichier Read Me :

- Associer aux données un fichier README
- Décrit les données pour les rendre compréhensibles et réutilisables.
- On y retrouve les instructions d'interprétation et de réutilisation des données.

### Exemple concret ?

[Fichier Read Me Cornell University](#)

[Guidelines for creating a README file](#)

### 3. Stockage et sauvegarde des données



La partie stockage du PGD vise à répondre aux enjeux liés à la sauvegarde et au partage des données au sein de l'équipe projet.



Elle vise à démontrer que des méthodes sont utilisées pour prévenir des **risques** de perte de données.



L'accès aux données est également contrôlé.

## 3. Stockage et sauvegarde des données

### **Sécurité des données :**

Récupération des données en cas d'incident, ne pas perdre de données en cours de projet.

Gestion des risques liés aux données, se fixer des règles pour limiter les risques.

### **Protection des données**

Contrôle des accès aux données.

Politique et interlocuteurs pour protéger les données.

Gagner du temps et de l'efficacité lors d'un travail collaboratif.

## 3. Stockage et sauvegarde des données

### S'interroger sur :

**Lieux de stockage** : répartition des données entre partenaires, supports physiques utilisés...

**Politique de sauvegarde** : processus utilisé, fréquence des sauvegardes, contrôle de sécurité...

### Quelques conseils :

Réaliser des sauvegardes régulières de vos données sur plusieurs supports sécurisés en appliquant dans l'idéal **la règle du 3-2-1** :

- garder 3 exemplaires des données,
- sur 2 supports ou technologies différents,
- dont 1 sauvegarde en ligne avec synchronisation.

Attention pas de Dropbox, Google Drive, One drive, etc.

## 3. Stockage et sauvegarde des données

### Quelques conseils :

**A chaque point d'étape** du projet, sélectionner les **données à sauvegarder**, à supprimer.

Définir leur durée de conservation

### Gérer les versions :

- Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement.
- Permettra de revenir à une version antérieure si besoin.



## 4.Exigences légales et éthiques

### Les données concernées :

- Données à caractère personnel
- Données sensibles
- Données médicales
- Données relevant du secret statistique
- Données d'entreprises

Concerne **l'ensemble des traitements** liés à ce type de données :

Réutilisation, collecte, conservation, diffusion...

## 4.Exigences légales et éthiques

### On s'interroge sur :

Si des données à caractère personnel sont traitées :

- Comment le respect des dispositions de la législation sur les données et sur la sécurité des données sera-t-il assuré ?
- 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 ?
- Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

## 4.Exigences légales et éthiques

### Focus :

#### **Données à caractère personnel :**

Veillez à ce que les lois sur la protection des données soient appliquées (RGPD)  
*anonymisation des données / pseudonymisation / cryptage*

Consultés les interlocuteurs dans ce domaine : *DPO, CNIL, comité d'éthique...*

Questions juridiques : titularité des droits, propriété intellectuelle...

Propriété des données : formalisme envisagé.

Réutilisation des données : licence d'accès et de réutilisation, restrictions envisagées...

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

### La conservation :

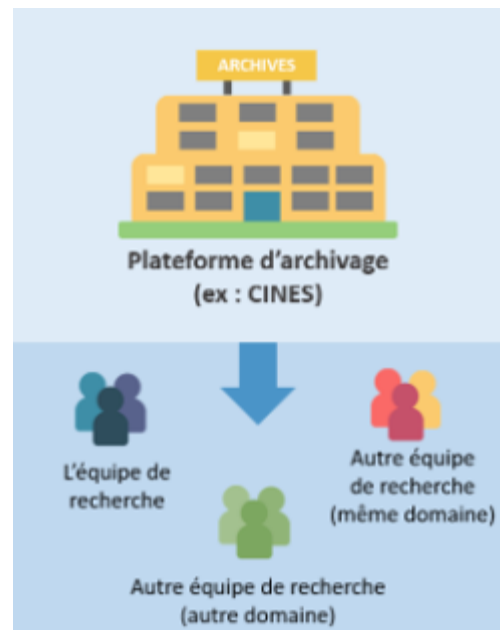
La conservation doit être différenciée du stockage.

En menant une réflexion sur la conservation, c'est sur l'après projet qu'on s'interroge et qu'on dépasse alors la durée de vie du projet.

### Pour aller plus loin :

[Stockage et archivage : fiche synthétique.](#)

[Stockage, partage et archivage quelles différences.](#)



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

### Le partage des données :

- S'interroger sur la valeur de vos travaux sans vos données ?
- Lesquelles de vos données ont le plus de valeur ? Pour les autres ? Pour quoi faire ?
- A qui appartiennent vos données ?
- Quelles données restent au laboratoire ? Que conservez-vous ?
- Vos données sont-elles soumises à des obligations (suppression ou conservation) ?
- Et après vous ?

Avoir des **données FAIR permet de faciliter** leur réutilisation...

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

### Partage des données :

- Quelles données allez-vous partager ? Y a-t-il un embargo ?
- Lesquelles ne seront pas ouvertes ? Pour quelles raisons ?

### Sélection/ conservation

- Quelles données conservées ? Pourquoi ?
- Quelles données doivent être supprimées ? Quelles sont les raisons ?
- Comment vos données pourront-elles être réutilisées ?

### Format des données (ouverts/fermés)

Des données faciles à trouver :

*Attribuer un identifiant pérenne (doi, ARK, PURL...)  
Utiliser un entrepôt de confiance.*

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

### Focus comment diffuser mes données ?

A l'aide d'un entrepôt de données généralistes :

*Zenodo, Dryad, Figshare*

zenodo



Dans des entrepôts de données spécialisés (*GenBank, GBIF, Software Heritage, Nakala, beQuali, progedo...*)

Recherche Data Gouv : l'entrepôt national des données de la recherche ouvert depuis juillet 2022.



RÉPUBLIQUE  
FRANÇAISE

Liberté  
Égalité  
Fraternité

recherche.data.gouv.fr

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

### Focus comment trouver un entrepôt fiable ?

Voir Re3data: <https://www.re3data.org/>

Voir CatOpidor: <https://cat.opidor.fr/index.php/Conservation>





## 6. Responsabilités et ressources

### **Responsabilité et ressources :**

**Décrire les rôles et responsabilités :** *Partenaires, noms des personnes selon les étapes du cycle de vie des données*

**Le responsable du plan de gestion de données :** *mise en œuvre, mise à jour, coordination.*

Quelles sont les ressources qui seront dédiées à la gestion des données et permettront de s'assurer quelles sont FAIR ?

**Temps de travail :** Qui ? Pour quelles tâches ?

**Frais de stockage :** coût, répartition entre partenaires

**Coûts de mise à disposition des données :** entrepôts, anonymisation, préparation des données.

## 7. Quelles aides pour la rédaction ?

### Quelles aides pour la rédaction ?



Outil collaboratif, gratuit et recommandé par l'ANR.

Propose différents modèles de PGD à compléter (établissements, structures, financeurs). <https://dmp.opidor.fr/>

Choisir le modèle qui correspond à votre financeur.

Facilite la rédaction du PGD grâce à des guides et exemples personnalisés.

Développement de nouveaux modèles de DMP plus structurés, « machine actionable ».

Pas conçu pour servir d'entrepôt de PGD, ni de données



Pour aller plus loin...

**Sur le cycle de vie de la donnée :**

<https://dorum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche/>

**Sur les principes FAIR :**

<https://dorum.fr/enjeux-benefices/principes-fair/>

**Sur les métadonnées :**

<https://dorum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche/>

**Sur la question des standards de métadonnées :**

<https://dorum.fr/metadonnees-standards-formats/standard-metadonnees/>

Pour aller plus loin...

**Sur le contexte national et européen de la science ouverte :**

**La politique Science Ouverte de l'ANR :**

<https://anr.fr/fr/lanr/engagements/la-science-ouverte/>

**Deuxième Plan national pour la science ouverte 2021-2024:**

<https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

**La science ouverte dans les projets Horizon Europe, un critère d'excellence :**

[https://ec.europa.eu/info/funding-tenders/opportunities/docs/20212027/horizon/other/events/20210421/open-science\\_en.pptx](https://ec.europa.eu/info/funding-tenders/opportunities/docs/20212027/horizon/other/events/20210421/open-science_en.pptx)



Le Département d'appui à la recherche du SCD (DAREC).

## Contactez-nous !

Pour toute question ou demande d'accompagnement, n'hésitez pas :

- ❖ Thèses, mémoires et HDR.
- ❖ Archive ouverte, HAL, open access
- ❖ Charte signature unique, bibliométrie
- ❖ Identifiants-chercheurs, identité numérique
- ❖ Gestion des données de la recherche
- ❖ Relecture de PGD

[servicechercheurs-bu@uphf.fr](mailto:servicechercheurs-bu@uphf.fr)

Merci pour votre attention