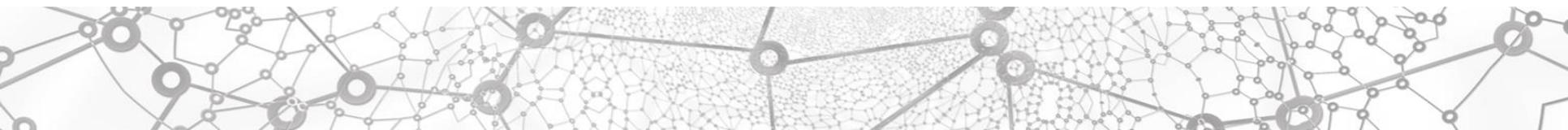


# Découverte du plan de gestion de données

Formation du mardi 15 mars 2021

Elise Lehoux & Benjamin Rullier

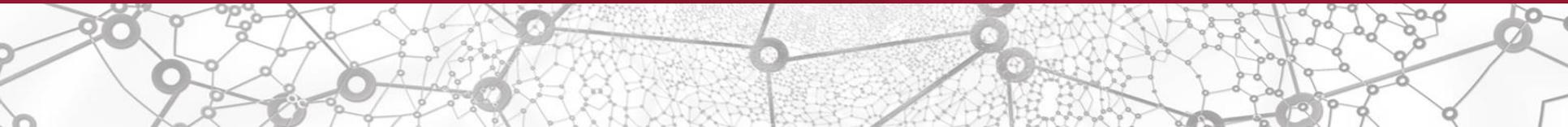
- 1. Les principes d'un Plan de gestion de données**  
(PGD ou *Data management plan* – DMP)
- 2. Les différentes parties d'un PGD**
- 3. Outils de rédaction d'un PGD**
- 4. Questions**



# Les principes d'un Plan de gestion de données



Source : <https://openbioeco.hypotheses.org/287>





## Les engagements européens et internationaux





MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR,  
DE LA RECHERCHE  
ET DE L'INNOVATION

Liberté  
Égalité  
Fraternité



7 oct. 2016

**Loi pour une République numérique – article 30**

○ Ouverture des données de la recherche « achevées » assimilées aux données administratives

4 juil. 2018  
**Plan national pour la science ouverte (PNSO)**

Rend obligatoire l'accès ouvert pour les publications et les données de recherches financées sur projets

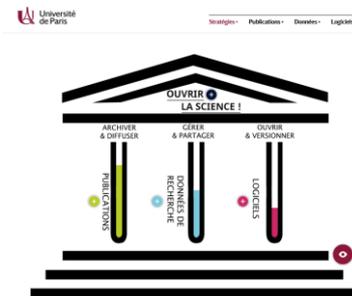
sept. 2019  
**ANR – ouverture des données de la recherche**

Rend obligatoire le Plan de gestion de données (DMP) pour les projets financés  
Dépôt des publications dans HAL

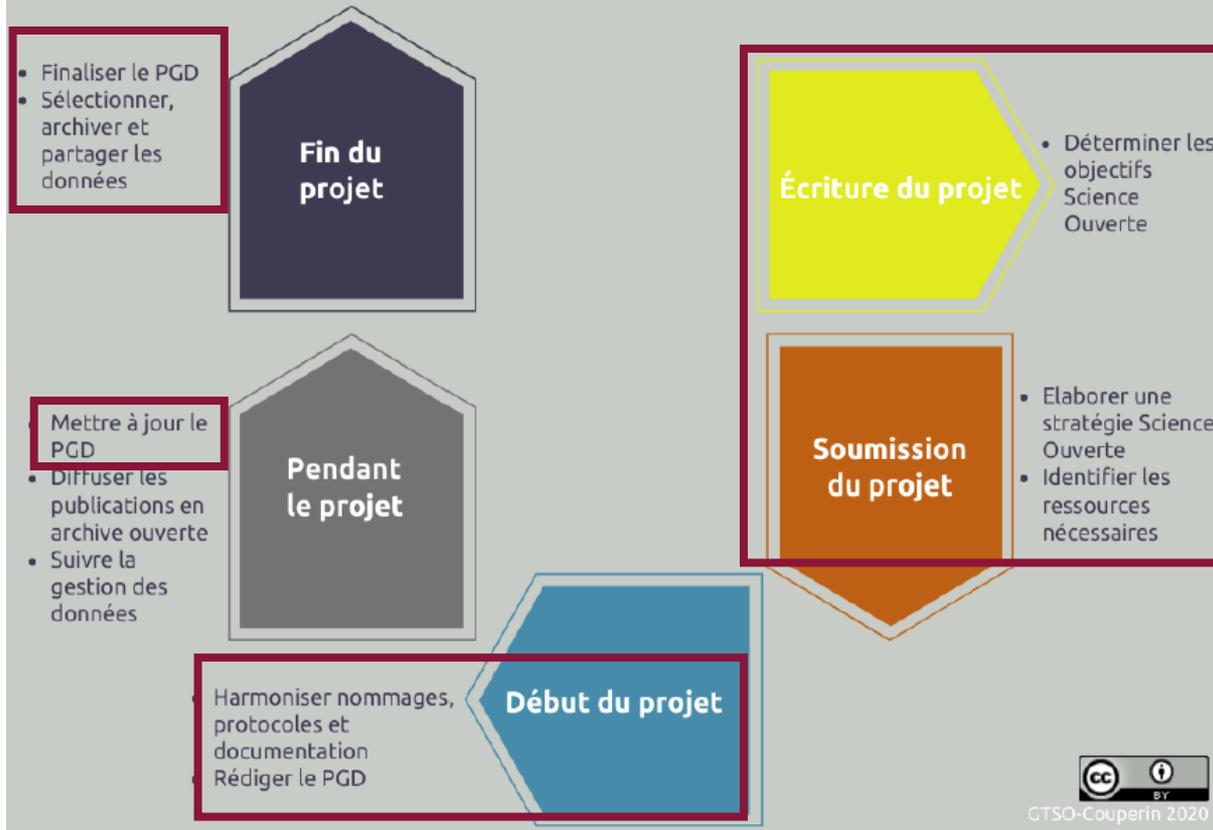
2021  
**Plan national pour la science ouverte (PNSO 2)**

2021  
**Feuille de route pour la stratégie Science ouverte d'UP**

... 2022  
**Entrepôt français pour les données simples Recherche Data Gouv**



# La Science Ouverte dans les projets



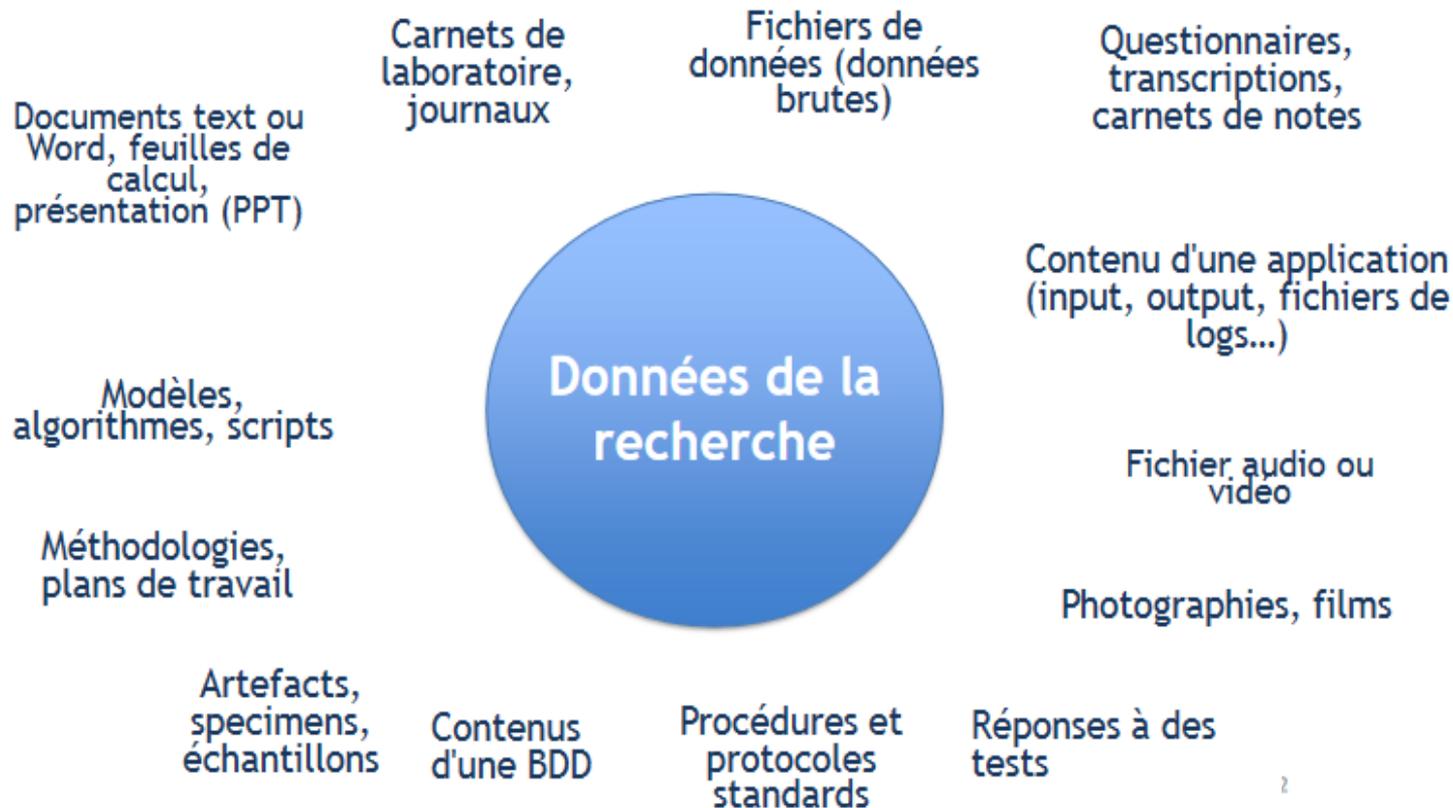
→ Réfléchir à **son plan de dissémination des résultats** (données et publications) dès le montage du projet

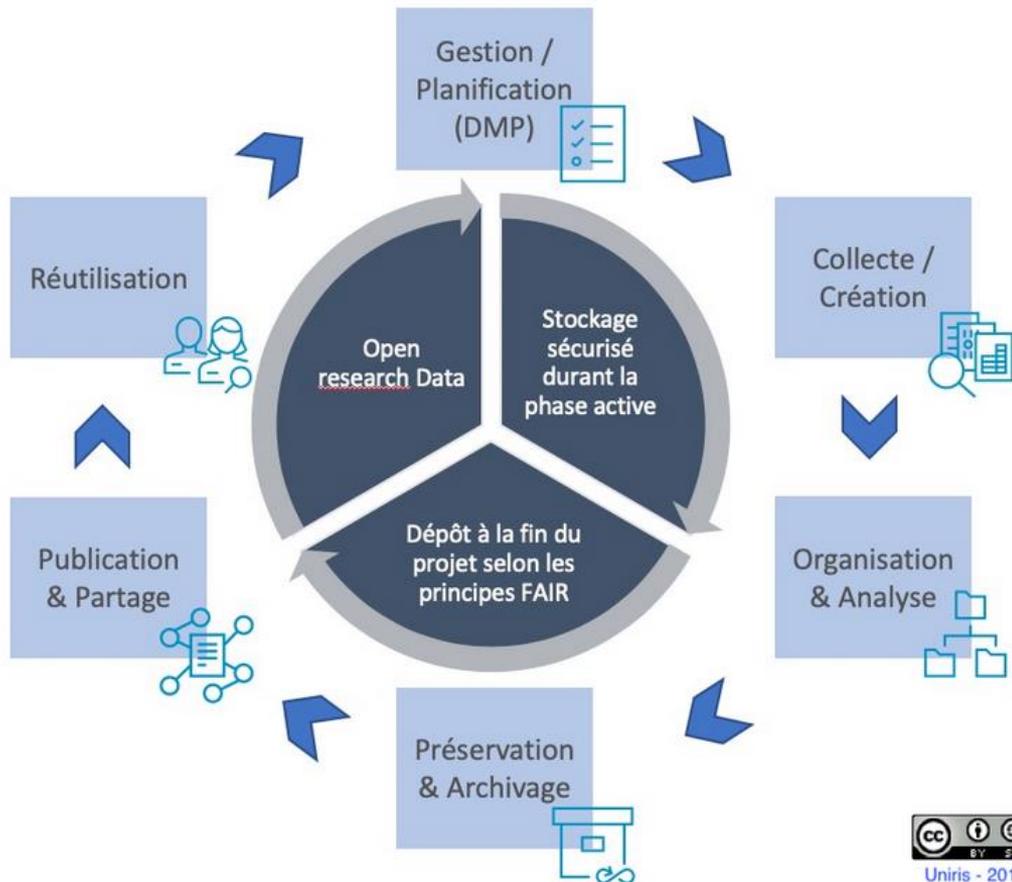
→ **Identifier les compétences Science Ouverte disponibles** au sein de l'équipe ou du consortium, **les compétences manquantes** et comment elles pourront être couvertes: formations, services supports de l'institution de rattachement, ressources extérieures, etc.

→ **Répartir les rôles et responsabilités au sein de son équipe** pour les activités qui relèvent de la gestion des données (tâches & livrables)

**Document synthétique et formel précisant la manière dont les données seront produites, traitées, décrites, partagées, protégées et conservées au cours et à l'issue du projet.**

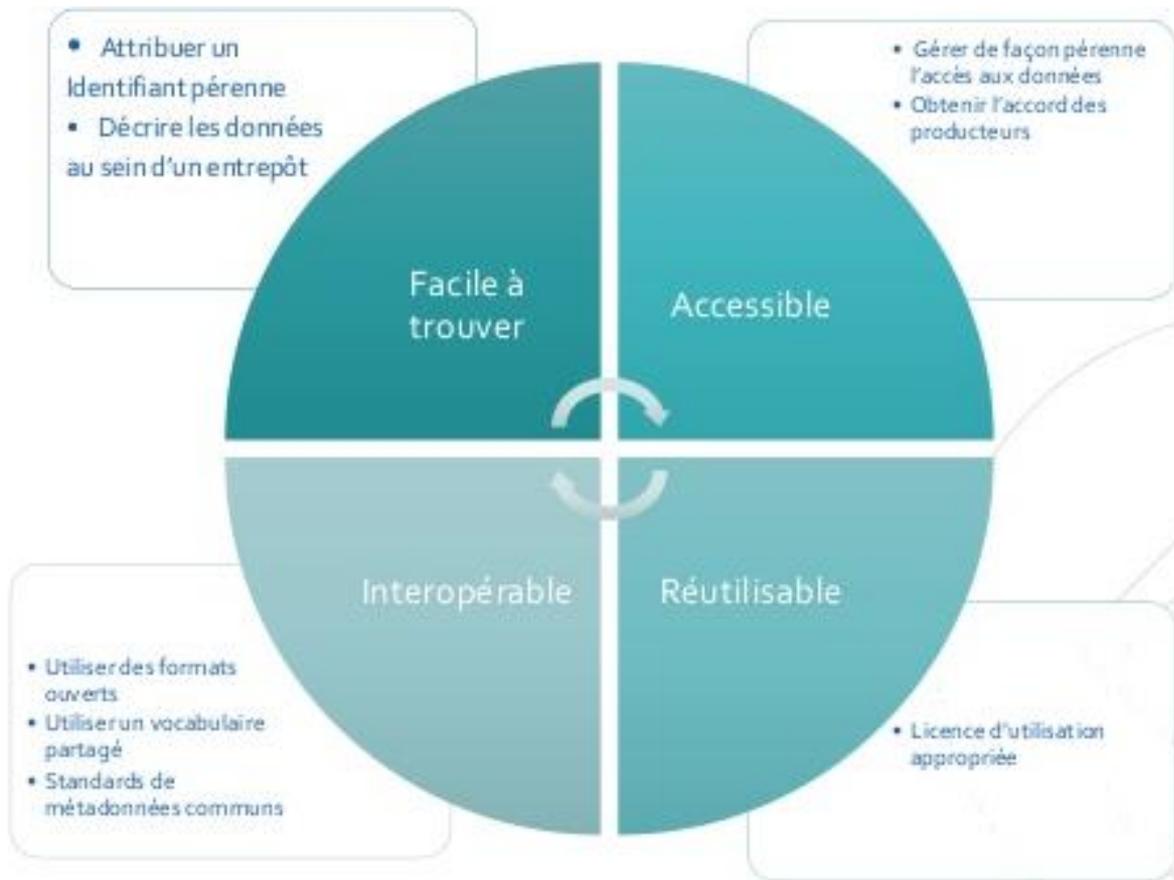
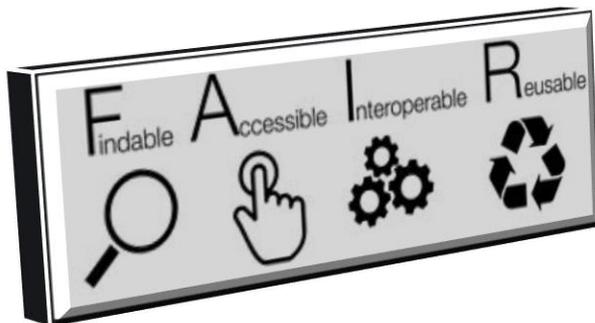
- **Décrit le cycle de vie** des données produites ou collectées au cours du projet de recherche.
- **Permet d'anticiper les questions de gestion** qui surviennent au cours d'une recherche et les conditions d'une diffusion et d'une conservation futures des données.
- **Itératif : il évolue et est actualisé** au cours du projet de recherche
- **Différents modèles possibles :**
  - Financeurs : ANR, Horizon 2020, ERC
  - Institutionnels
  - Disciplinaires ou thématiques





Pendant un projet de recherche et après son achèvement :

- Rassembler
- Organiser,
- Décrire,
- Stocker,
- Sauvegarder,
- Préserver,
- Partager vos données



|  | Horizon Europe  | ANR   |
|--|---|---|
| Obligatoire  | Oui   | Oui   |
| Courte description de la politique de gestion des données dès le dépôt de la proposition | Obligatoire   | Pas obligatoire mais vivement conseillé   |
| Organisation des lots de travail (work packages ou WP) et des tâches                     | <ul style="list-style-type: none"><li>• Dans le WP 1 « Gestion du projet » ou dans le WP consacré à la diffusion des résultats (s'il existe)</li><li>• Une ou deux tâches dédiées</li></ul> | <ul style="list-style-type: none"><li>• Dans le WP 1 « Gestion du projet »</li><li>• Une ou deux tâches dédiées</li></ul> |
| Première version   | Dans les 6 premiers du projet   | Dans les 6 premiers du projet   |
| Version à mi-projet  | Pas obligatoire mais vivement conseillée  | Obligatoire pour les projets > 30 mois<br>Autrement : pas obligatoire mais vivement conseillée                            |
| DMP final à la fin du projet   | Obligatoire   | Obligatoire   |

Certaines recherches scientifiques traitent des données à caractère personnel. A ce titre, la recherche met en œuvre des traitements de données personnelles qui doivent être inscrits au sein du registre des activités de l'établissement, tenu par le Délégué à la protection des données. En outre, il est nécessaire de s'assurer de la conformité de la recherche à la réglementation sur la protection des données (RGPD), pour laquelle le DPD peut vous accompagner.

| CHERCHEUR  | DPO  |
|--|--|
| <b>Etape 1 : Transmettre à <a href="mailto:dpo@u-paris.fr">dpo@u-paris.fr</a></b>  | <b>Etape 2 : Envoi documents conformité à remplir aux chercheurs + analyse</b>               |
| <ul style="list-style-type: none"><li>- Protocole de recherche</li><li>- Notice d'information</li><li>- Formulaire de consentement</li><li>- Avis du CCP/Comité d'éthique</li><li>- Plan de gestion de données (si possible)</li></ul> | Analyse de conformité des documents transmis par le chercheur (mails + entretiens tel/visio) |
| <b>Etape 3 : Renseigner et adresser docs au DPD</b>  | <b>Etape 4 : Inscription recherche au registre</b>   |
| <ul style="list-style-type: none"><li>- Feuille de registre</li><li>- Document de conformité MR à remplir (le cas échéant)</li></ul>   | Enregistrement au sein du registre de traitements d'U-Paris                                  |

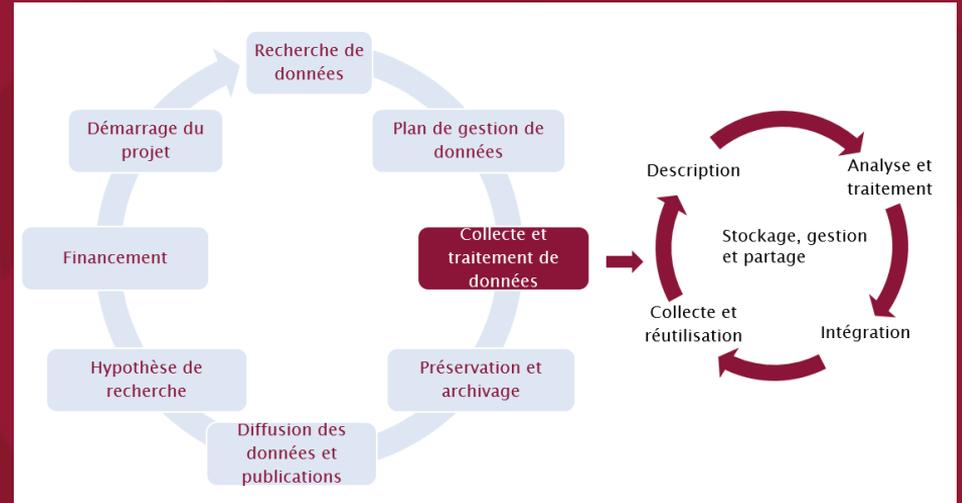
## Règles d'or de la protection des données

- > **Collecter uniquement les données réellement nécessaires à la recherche**  
(ex. ~~01/11/2001~~ 11/2001/classe d'âge)
- > **Etre transparent**  
(Notice d'information et/ou formulaire de consentement)
- > **Attention aux droits des personnes**  
(mentions d'information RGPD participants et personnel de la recherche/Procédure de gestion des droits/coordonnées de contact)
- > **Sécuriser les données**  
(Chiffrement ou conteneurs sécurisés/transferts de fichier sécurisés « File Sender », gestion des droits d'accès/antivirus)

## Acteurs de la protection des données personnelles

- > **Le responsable du traitement (RT) :**  
(Détermine la finalité et les moyens nécessaires  
Présidente d'Université de Paris (RT) ou si UMR, le Directeur d'unité est le RT (Note CPU-CNRS du 04-09-2017))
- > **Personne concernée : personne physique**  
(participants à la recherche/personnel de la recherche)
- > **Le Délégué à la protection des données (DPD)**
- > **Le sous-traitant de données personnelles**  
(éditeur de logiciel/hébergeur de données, retranscription,...)

# Les différentes parties d'un PGD



## Le modèle de la CE pour Horizon Europe

1. **Data Summary**
2. **FAIR Data**
  1. **Making data findable, including provision for metadata**
  2. **Making data accessible**
  3. **Making data interoperable**
  4. **Increase data re-use**
3. **Other research output**
4. **Allocation of ressources**
5. **Data security**
6. **Ethics**
7. **Other issues**

## ANR

1. Description des données et collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le processus de recherche
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
6. Responsabilités et ressources en matière de gestion des données

# 1. Description des données / collecte et réutilisation de données existantes

## 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Selon le projet, les données de la recherche peuvent être :

- **produites** : ce sont les données créées, élaborées, générées lors d'activités de recherche (observations, mesures...)
- **collectées** : ce sont des données déjà existantes (corpus, archives...) qui sont utilisées pour le projet
- **produites et collectées**

→ Expliquer la méthodologie de collecte et/ou l'origine des données

## 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données utilisées devront être documentées :

- **type de données** collectées ou produites
- **format** des données : en justifiant si un choix se présente (format libre ou propriétaire) et distinguant si possible le format pour la diffusion de celui pour la conservation.
- **volume de données produit** (approximatif en début de projet).

## 2. Documentation et qualité des données

### 2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?



Les métadonnées sont **des informations descriptives qui permettent de renseigner le contenu d'un jeu de données** :

- Description du contenu (ex : titre, auteur, date de création, format, etc.)
- Règles de gestion (ex : règles d'accès et de partage, méthode de collecte, etc.)
- Informations techniques (ex : résolution, dimension, poids, encodage, etc.)

#### **Adopter des standards de métadonnées disciplinaires :**

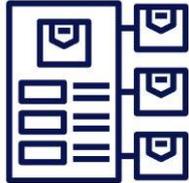
- Les jeux de données seront plus facilement repérables et réutilisables par les communautés et conformes aux principes FAIR (Findable, Accessible, Interoperable, Reusable)
- Listes de standards disciplinaires sur le site du [Digital Curation Center](#)

#### **Adapter un standard de métadonnées générique :**

- Si aucun standard n'est utilisé dans la discipline
- Exemple de standard générique : [Dublin Core](#)

## 2. Documentation et qualité des données

### 2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?



#### Définir des règles de gestion des données au cours du projet :

- partagées par l'équipe de recherche et formalisées dans un document pouvant être un fichier [ReadMe](#)
- précisant notamment :
  - Contacts au sein de l'équipe
  - Méthodologie de collecte, de contrôle qualité des données
  - Règles de nommage, d'organisation des données
  - Règles de partage et d'accès aux données au sein de l'équipe de recherche
  - Vocabulaire utilisé, etc.

#### Définir des règles d'organisation et de nommage des dossiers et fichiers

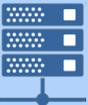
- facilite le partage et permet de retrouver rapidement les données
- Exemples de fiches pratiques :
  - [fiche pratique « Bien nommer ses dossiers et fichiers »](#)
  - [fiche pratique « Bien organiser son arborescence documentaire »](#)

### 2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Expliquer ici comment la qualité et la conformité de la collecte des données seront contrôlées et documentées : revue par les pairs, méthodes (calibration, répétition des échantillons, capture standardisée, etc.) ou les démarches qualité engagées par la structure (ex : certification ISO 9001).

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

### 3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

| Support de stockage   | Sécurité   | Accès   | Coût   | Remarque d'utilisation   |
|---|--|---|--|--|
| <br>Ordinateur professionnel | ★★☆☆☆<br>Sujet au piratage informatique, aux détériorations et pannes                              | ★☆☆☆☆<br>Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...) | ★★★★★<br>Pas de coût supplémentaire ou coût peu important                    | - Pour un stockage temporaire<br>- Nécessité de crypter les données confidentielles et sensibles   |
| <br>Support externe          | ★☆☆☆☆<br>- Sujet au vol, à la perte du support<br>- Durée de vie limitée (dégradation du matériel) | ★★★★★<br>Facilement transportable, il permet de transférer les données vers un autre ordinateur             | ★★★★★<br>Pas de coût supplémentaire ou coût peu important                    | - Pour un stockage temporaire<br>- Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles  |
| <br>Serveur institutionnel   | ★★★★★<br>Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)       | ★★☆☆☆<br>La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures   | ★★★☆☆<br>Coût assez important mais pas forcément répercuté sur l'utilisateur | - Pour un stockage plus pérenne<br>- Adapté pour le stockage de données sensibles et des versions « stables » de vos données<br>- Toutes les institutions ne proposent pas ce service      |
| <br>Serveur Cloud            | ★★☆☆☆<br>On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent           | ★★★★★<br>Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage            | ★★★☆☆<br>Payant à partir d'une certaine limite de stockage                   | - Pour un partage avec des personnes externes à l'institution<br>- Ne pas y mettre de données sensibles ou confidentielles<br>- Pas de contrôle sur la procédure de sauvegarde des données |



**Faites des sauvegardes régulièrement :**

- 3 sauvegardes
- 2 supports différents
- dont 1 à distance



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

### 3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Mesures de sécurité des données pouvant être mises en œuvre :

- Respect de la **charte informatique** et/ou de la **politique des systèmes d'information** (PSSI) de l'établissement.
- **Configuration et contrôle des droits d'accès** attribués aux membres de l'équipe de recherche : authentification, traçabilité des accès, contrôle physique par badge, etc.
- **Partage sécurisé à distance des fichiers** entre les membres : chiffrement des fichiers lors des envois, utilisation d'un canal sécurisé de transfert/partage des fichiers comme [FILESENDER](#) ou [France transfert](#), pour de gros volume de données non sensibles (jusqu'à 20 Giga).
- Mesures techniques spécifiques à la **sécurité des données les plus sensibles** : protection des fichiers par mot de passe, chiffrement des fichiers ou du disque dur, utilisation de conteneurs sécurisés, etc.
- Mesures de **restauration des données en cas de perte** de données (systèmes de sauvegardes pour restauration d'une précédente version des données, etc.).
- **Contrôle de l'intégrité des données** (« journal des événements sur les données » etc.)
- Pour les données existantes sur support papier ou pour les copies des données sur des supports amovibles (disques durs externes par ex), préciser de quelle manière ces supports sont conservés de manière sécurisée (dans un tiroir fermé à clef, au sein de l'établissement ou non, chiffrés ou protégés par mot de passe par ex.)

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

**4a. 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 ?**

Si votre projet, prévoit la collecte et l'utilisation de données à caractère personnel, **contactez le Délégué à la Protection des Données** (DPD) désigné pour l'UMR (DPD de l'une des tutelles) **avant toute opération de collecte de données**

- Le traitement de données à caractère personnel doit être **déclaré dans le registre des activités de traitement** tenu par le DPD
- Si la collecte des données repose sur le consentement des personnes, les **modalités de recueil et de gestion du consentement** au cours du projet de recherche doivent être précisées (possibilité de se rétracter et de sortir du champ de l'étude)
- Les **moyens pour informer les personnes concernées** du traitement de leurs données, ainsi que les **modalités d'exercice des droits des personnes concernées** doivent être détaillées
- Ainsi que, les **mesures techniques et organisationnelles pour garantir la confidentialité et la sécurité des données** (techniques de pseudonymisation, chiffrement des données, etc.)
- La **réalisation d'une analyse d'impact sur la protection des données (AIPD)** pour le traitement de données très sensibles, en lien avec le DPD, précisant les risques encourus et les mesures mises en œuvre pour protéger ces données

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

**4b. 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 ?**

La propriété des données est généralement fixée par l'**accord de consortium** définissant le rôle et les apports des partenaires du projet.

La négociation des termes de l'accord en amont du projet est essentielle car elle conditionne en partie les modalités de partage et de dissémination ultérieure des données et résultats.

S'il n'en existe pas, elle sera précisée dans le plan de gestion de données.

**4c. Comment les éventuelles questions éthiques seront-elles prises en compte, et les codes déontologiques respectés ?**

Si des questions d'éthique pourront se poser avec le projet de recherche (conflit d'intérêt, traitement de données « sensibles » du point de vue sociétal, etc.) : indiquez les éventuelles incidences de ces questions d'éthique sur la façon dont les données seront stockées et transférées, qui pourra les voir ou les utiliser et quelles durées de conservation leur seront appliquées.

Définissez les mesures qui seront mises en œuvre pour répondre à ces questions éthiques (application de codes de conduite, ou codes d'éthique et validation du projet de recherche par un comité d'éthique).

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

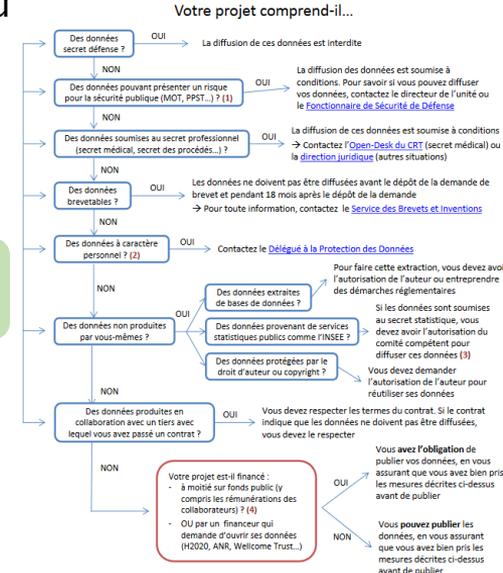
### 5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Le partage de jeux de données - s'il est possible - peut se faire par différentes voies :

- le **dépôt des jeux de données dans un entrepôt** (accès ouvert, restreint ou fermé, avec ou sans embargo)
- l'**indexation des métadonnées dans un catalogue**
- La **publication de *data papers*** associée au(x) jeu(x) de données, etc.

Le principe général guidant l'ouverture des données de la recherche est le suivant :  
**« Aussi ouvert que possible, aussi fermé que nécessaire »**

- Mes données sont-elles soumises à une obligation réglementaire, contractuelle ou légale ?
- Que dois-je faire et que puis-je faire avec mes données ?



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

### 5a. 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 ?

Le choix d'un entrepôt peut se faire par le biais de répertoires comme

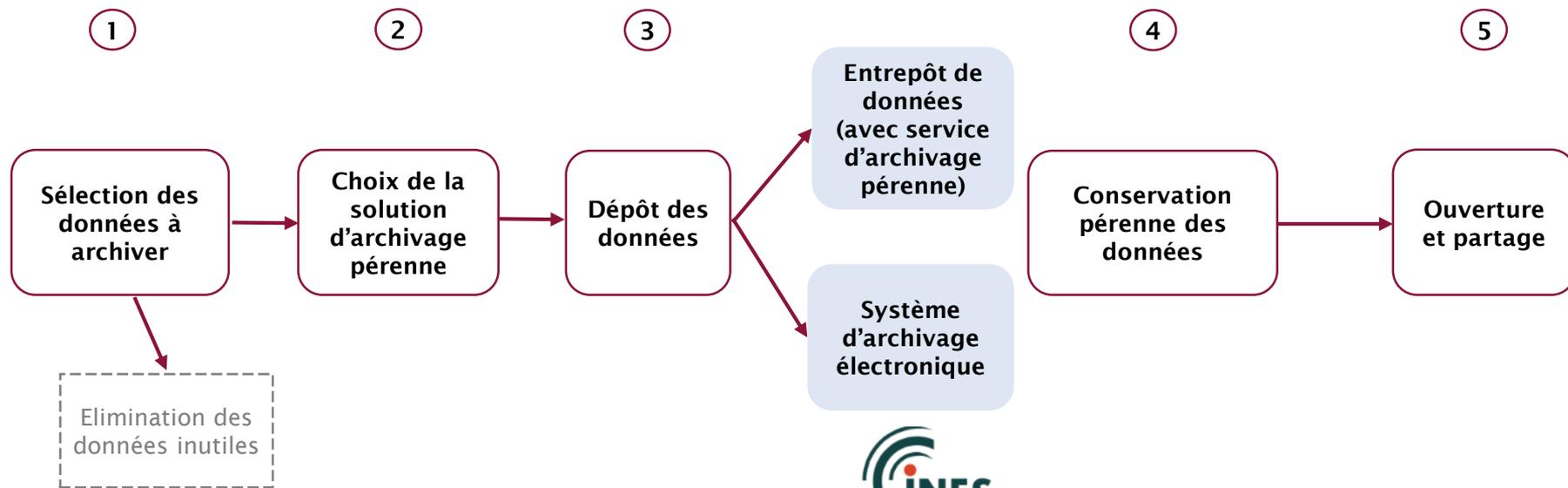


Et en suivant certaines recommandations, comme :

- choisir un entrepôt si possible spécialisé pour un type de données/sa discipline
- recommandé par une communauté
- durée de préservation des données
- qualité de préservation des données
- certification de l'entrepôt
- identifiants pérennes associés aux jeux de données
- précisions des métadonnées
- mode d'accès
- licences acceptées

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

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (p. ex. un entrepôt de données ou une archive) ?



par ex.  INES  
Centre Informatique National  
de l'Enseignement Supérieur

Le [référentiel de tri et de conservation des archives de la recherche](#) disponible sur le site de DoRANum peut vous aider à sélectionner les données à archiver

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

**6a. Qui (p. ex. rôle, position et institution de rattachement) sera responsable de la gestion des données (le gestionnaire des données) ?**

Désigner un responsable de la gestion des données qui sera entre autre, en charge de la mise à jour du PGD.

**6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?**

Trois sites peuvent être utiles pour évaluer les coûts :

UK Data Service Data management costing tool and checklist :

<https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>

L'Université d'Utrecht détaille les coûts induits par la gestion des données à toutes les étapes du cycle de vie :

<https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>

L'Université de Lausanne a développé un calculateur de coût induit par la gestion, le stockage et la publication des données : <https://costcalc.epfl.ch/>

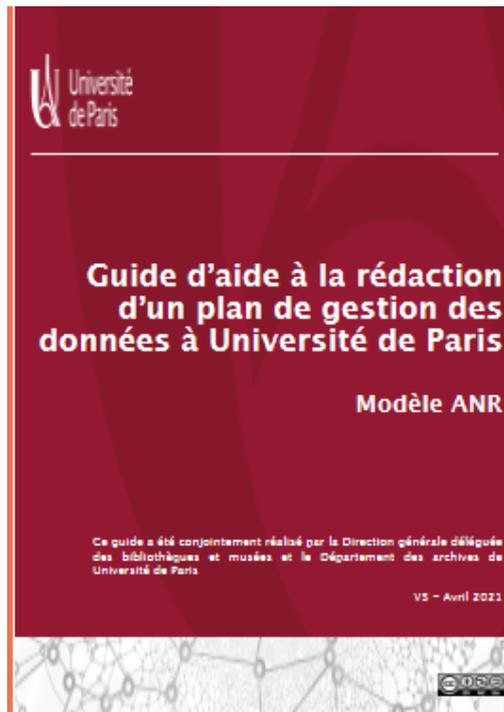
# Un outil d'aide à la rédaction : [DMP OPIDoR](#)



## Un outil de rédaction d'un PGD : DMP OPIDoR



## Un guide d'aide à la rédaction d'un PGD



## Un guide sur la Science ouverte dans les projets financés

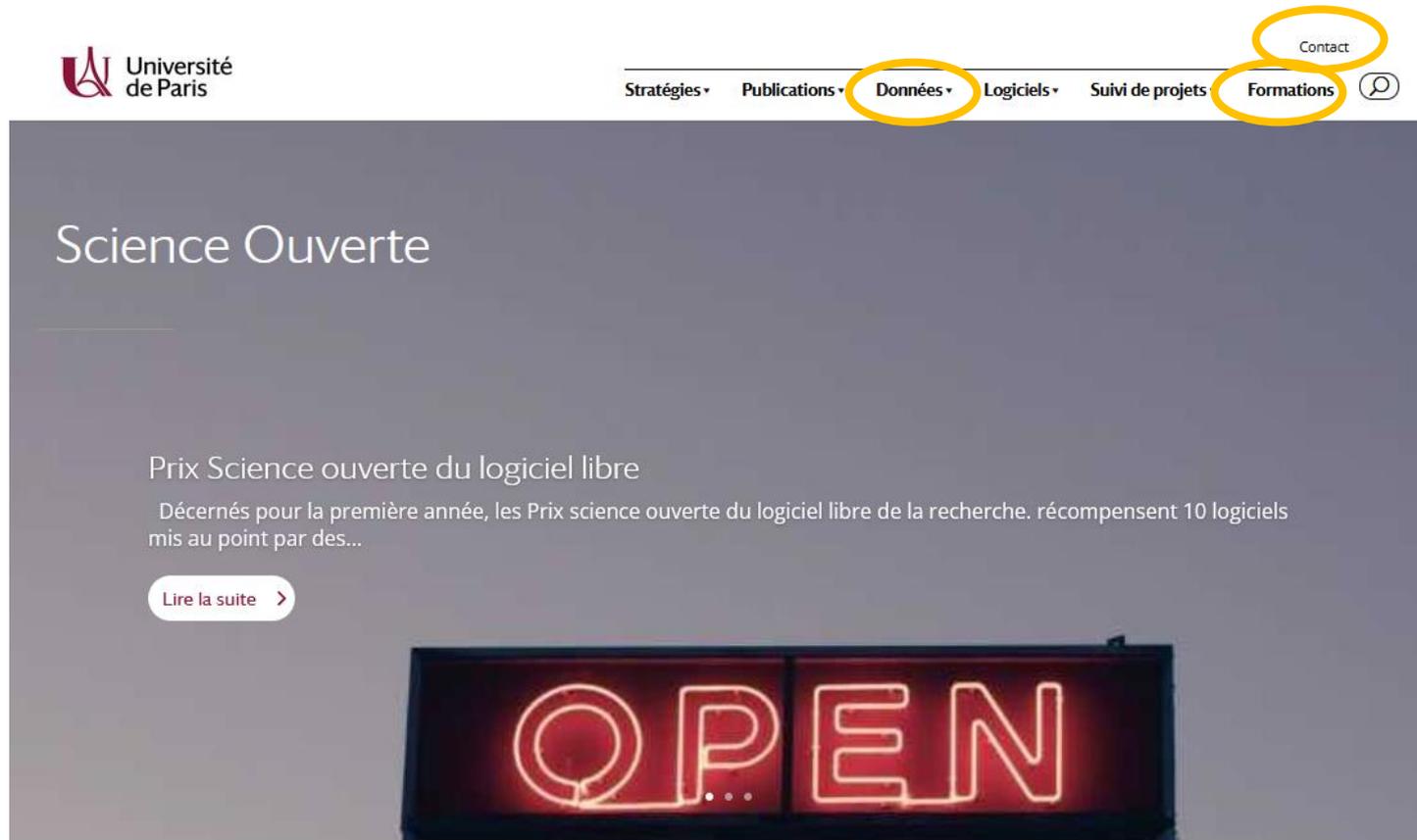
couperin.org  
Consortium Univ. des Établissements Universitaires et  
de Recherche pour l'Accès aux Publications Numériques

Améliorer son projet ANR grâce à la Science  
Ouverte

Guide produit par le Groupe Données du GTSO-Couperin  
Avril 2020  
Version 2  
DOI : [10.5281/zenodo.3769954](https://doi.org/10.5281/zenodo.3769954)

Une équipe à UP vous accompagne → [donnees.recherche.dbm@listes.u-paris.fr](mailto:donnees.recherche.dbm@listes.u-paris.fr)

## Pour retrouver toutes les informations sur la Science ouverte à UP



Universit  de Paris

Stratgies • Publications • Donn es • Logiciels • Suivi de projets • Formations • Contact

# Science Ouverte

### Prix Science ouverte du logiciel libre

D cern s pour la premi re ann e, les Prix science ouverte du logiciel libre de la recherche. r compensent 10 logiciels mis au point par des...

[Lire la suite >](#)

OPEN

**Contacts :**

[elise.lehoux@u-paris.fr](mailto:elise.lehoux@u-paris.fr)

[benjamin.rullier@u-paris.fr](mailto:benjamin.rullier@u-paris.fr)

Et/ou [donnees.recherche.dbm@listes.u-paris.fr](mailto:donnees.recherche.dbm@listes.u-paris.fr)

Le diaporama vous sera envoyé dans les prochains jours ;  
tous les liens, photos, références sont [cliquables](#).

Merci pour votre attention !