



Plan de Gestion des Données (PGD) — Data Management Plan (DMP)





C'est :

- une aide concrète à la gestion des données durant et après la phase de recherche
- un outil normalisé, évolutif tout au long du projet, et de plus en plus automatisé
- un livrable du projet dans le cas d'un financement



PGD / DMP (Data Management Plan) = document bureautique



Votre expérience ?

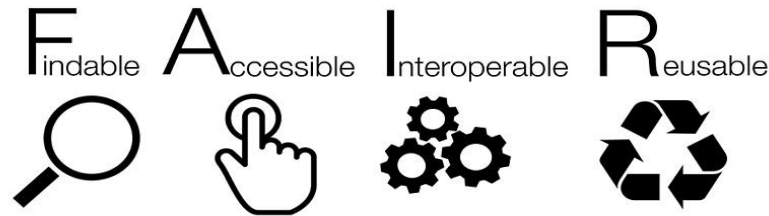
Si votre thèse s'inscrit dans un projet ANR, européen, avez vous consulté le PGD du projet ?



Le Plan de Gestion des Données **permet de réfléchir à la gestion des données au cours de votre thèse :**

- **Quelles données vont être obtenues**
- **Comment vont-elles être utilisées**
- **Comment elles vont être préservées**
- **Comment elles vont être valorisées**
- **Comment assurer le financement des ressources nécessaires ?**

Pour répondre aux principes FAIR de l'open data



Findable / *Trouvable*

Données **faciles à trouver**.

- possédant un identifiant unique et pérenne
- décrites par des métadonnées riches
- enregistrées ou indexées dans une source interrogeable

Interoperable / *Interopérable*

Facile à combiner avec d'autres jeux de données, par les humains **et** les systèmes informatiques

- formats libres et ouverts
- mise à disposition du code source si le logiciel de traitement existe
- standards de métadonnées et vocabulaire standardisés

Accessible / *Accessible*

Données ou au moins méta-données facilement accessibles.

- entrepôt de confiance, pérenne, certifié
- définir les conditions d'accès et la licence de diffusion
- si embargo ou accès restreint : méta-données accessibles

Reusable / *Réutilisable*

Prêtes à être **réutilisables** pour une future recherche y compris via des méthodes informatiques

Pourquoi faire un PGD pour sa thèse ?



- **Pendant la thèse :**

- pour vous : un repère, un outil à faire évoluer, une liste des questions à résoudre
- pour votre directeur de thèse : permettre les échanges avec lui, les points de discussion
- pour l'équipe dans laquelle vous travaillez : faire comprendre ce que vous faites
- pour vos partenaires éventuels : rendre compte

N'oubliez pas que votre thèse dure 3 ans

- **Après votre thèse :**

- pour ceux qui prendront votre suite
- pour les labos dans lesquels vous souhaitez travailler après votre thèse
- pour préparer une demande de financement de type ANR pour continuer vos travaux



- Vous devez expliquer à votre directeur de thèse comment vous allez gérer vos données ...
- Pour cela, nous vous donnons une liste de questions
- Consignes : à remplir pour la session suivante
- **Attention :**
 - il n'y a pas de mauvaise réponses
 - pas de souci si vous n'arrivez pas à répondre à toutes les questions
- Lors de la deuxième session : on fera une restitution afin de vous aider à avancer

Si vous avez des questions d'ici la deuxième session, vous pouvez contacter : sos-data@univ-grenoble-alpes.fr

À quelles questions répondre ?



1. Description des données et collecte ou réutilisation de données existantes

En cas de collecte de données, d'où viennent-elles (par ex, d'un entrepôt, autre ...) ?

.....

.....

.....

En cas de production de données, quelles méthodes seront appliquées pour les produire ? Quels logiciels, outils, seront utilisés ? (par exemple, logiciels d'utilisation d'un instrument, etc.)

.....

.....

.....

Quelle est la nature des données produites ou réutilisées (par ex, RMN, capteur météo, statistiques, données d'enquêtes, données médicales) ?

.....

.....

Quel est le type des données (par exemple image, audio, texte, etc.) ?

.....

.....

.....

À quelles questions répondre ?



1. Description des données et collecte ou réutilisation de données existantes

Quel est leur format (par ex, csv, HDF5, ..) ? Format ouvert ou fermé ? Si fermé, pour quelle raison ?

.....

.....

.....

Des outils ou des logiciels seront-ils nécessaires pour le utiliser ? (par ex, librairies Python, R,..)

.....

.....

.....

Quelle est la volumétrie prévisionnelle des données ? (par ex, 10 Go, 1 To ..)

.....

.....

.....

À quelles questions répondre ?



2. Documentation et qualité des données

Comment les données seront-elles organisées pendant le projet ? (préciser la structuration des dossiers, la convention de nommage des fichiers, la gestion de version)

.....

.....

.....

Quels éléments seront décrits via les métadonnées (par ex, procédures et méthodes d'analyse, les paramétrages et les définitions de variables, les unités de mesure, le titre, l'auteur, date d'enregistrement, géolocalisation, variables, mots-clés, résumé, etc)

.....

.....

.....

Quels seront les standards ou schémas de métadonnées utilisés pour renseigner les métadonnées ? (Par ex Dublin Core, Datacite Metadata, DCI, EML etc.)

.....

.....

.....

À quelles questions répondre ?



2. Documentation et qualité des données

Une procédure de contrôle qualité des données est-elle envisagée ? (par ex, répétition expérience, relecture des données d'enquête)

.....

.....

.....

Dans vos laboratoires, y-a-t-il une personne responsable de la qualité des données et de leur documentation ? Avez-vous discuté avec elle ?

.....

.....

.....

À quelles questions répondre ?



3. Stockage et sauvegarde pendant le processus de recherche

Où et comment seront stockées physiquement et sauvegardées les données pendant votre thèse ? (sur quel support, sur quel type d'hébergement : par ex, sur un disque dur, sur le serveur de votre labo, etc.)

.....

.....

.....

Quelle procédure de sauvegarde sera-t-elle mise en place (manuelle, automatisée) ? A quelle fréquence (par exemple, quotidienne, hebdomadaire) ?

.....

.....

.....

Si vous travaillez avec des partenaires (entreprise, etc.), des droits d'accès seront-ils nécessaires pour accéder à vos données pendant votre thèse ? Si oui, comment sont gérés les accès ?

.....

.....

.....

Comment les données seront-elles récupérées en cas d'incident ?

.....

.....

.....

À quelles questions répondre ?



4. Exigences légales et éthiques, codes de conduite (données à caractère personnel)

Quelles questions éthiques sont soulevées par votre thèse ?

.....

.....

.....

Quel code institutionnel, nationale ou international d'éthique sera appliqué ? Quelles mesures seront mises en place pour respecter ce code ?

.....

.....

.....

Si vous traitez des données personnelles :

Quel démarche avez-vous faite ? (par exemple solliciter le DPD, etc.)

.....

.....

.....

Quelles seront les mesures prises et normes choisies pour garantir la confidentialité des données personnelles (pseudonymisation ou anonymisation, chiffrement, procédure d'accès...) ?

.....

.....

.....

À quelles questions répondre ?



4. Exigences légales et éthiques, codes de conduite (données à caractère personnel)

Quelles seront les modalités pour informer les personnes et quels formulaires de consentement éclairé, d'autorisation de diffusion seront utilisés ?

.....
.....
.....

Si vous traitez des données sensibles (données médicales, liées aux appartenances politiques, syndicales, religieuses, données pouvant induire des discriminations) :

Qui sera chargé d'assurer les formalités de déclaration de données sensibles ?

Quelles seront les mesures prises et normes choisies pour garantir la confidentialité des données personnelles (pseudonymisation ou anonymisation, chiffrement, procédure d'accès...) ?

.....
.....

Quelles seront les modalités pour informer les personnes et quels formulaires de consentement éclairé, d'autorisation de diffusion seront utilisés ?

À quelles questions répondre ?



4. Exigences légales et éthiques, codes de conduite (données à caractère personnel)

À qui appartiennent les données que vous produisez (par ex, votre laboratoire, une entreprise, etc.) ?

.....

.....

Si vous diffusez vos données, sous quelle licence (par ex, Creative Commons, etc.) ?

.....

.....

À quelles questions répondre ?



5. Partage des données et conservation à long terme

Pensez-vous que les données utilisées dans votre thèse pourront être utiles à autres chercheurs ? Si oui, pourquoi ?

.....
.....
.....

Envisagez-vous de diffuser les données utilisées dans votre thèse ?

Si oui, où ? Par ex, entrepôts de données, datapaper, dans les annexes ...)

.....
.....
.....

Si oui, selon quelles modalités ? Par ex, immédiatement après la soutenance, avec un délai, en lien avec une publication

.....
.....
.....

Si oui, qui aura accès aux données après votre thèse ? Accès ouvert ou restreint ?

.....
.....
.....

À quelles questions répondre ?



5. Partage des données et conservation à long terme

Quels jeux de données de votre thèse seront librement accessibles ? (totalité des jeux de données ou sélection, données brutes ou données analysées, données liées à une publication)

.....

.....

.....

Quelles données seront conservées sur le long terme ?

Où seront-elles conservées ? (par ex, sur le serveur de votre laboratoire, ...)

.....

.....

.....

À quelles questions répondre ?



6. Responsabilités et ressources en matière de gestion des données

Qui sera responsable de la gestion de vos données après votre thèse ? (Par ex, ingénieur du laboratoire, directeur de thèse, chef d'équipe du laboratoire, chef de projet, etc.)

.....

.....

.....