



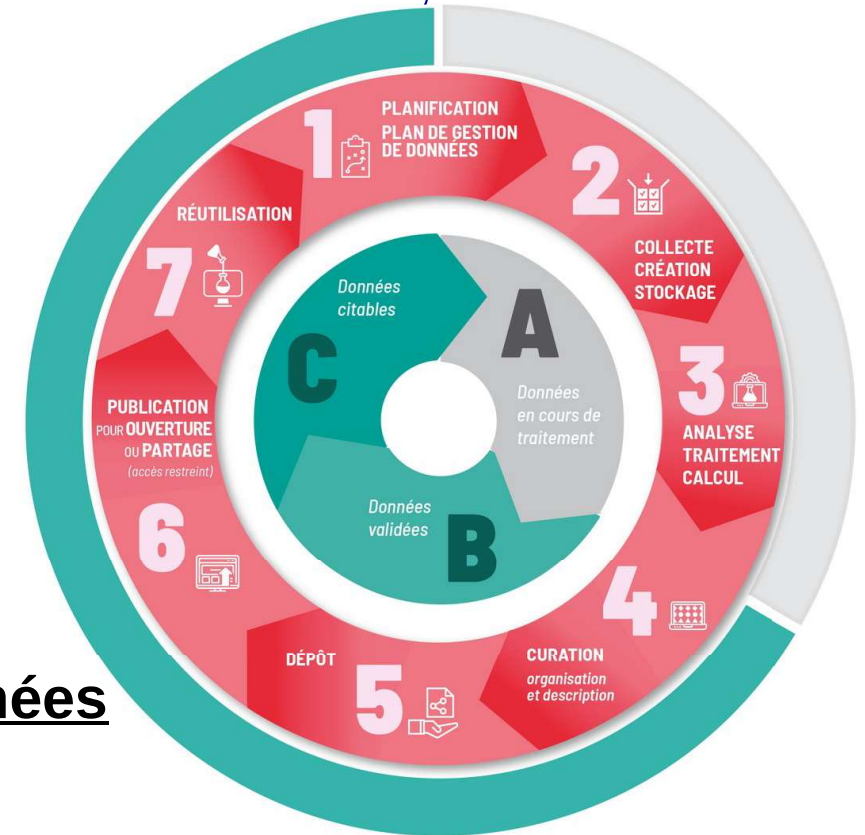
Plan de Gestion des Données (PGD) — Data Management Plan (DMP)





Cycle de vie des données, Recherche Data Gov

- Document formel qui décrit la façon dont les données seront :
 - **obtenues, organisées**
 - **stockées, sécurisée**
 - **traitées,**
 - **préservées, partagées** au cours et à l'issue d'un projet.



C'est :

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

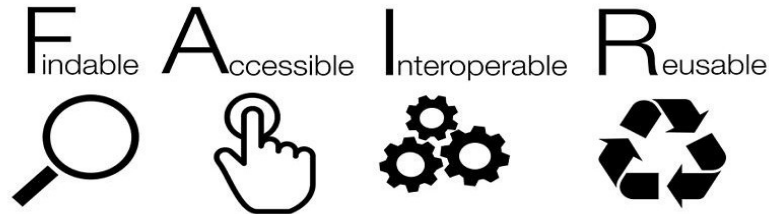
PGD / DMP (Data Management Plan) = document bureautique



Le PGD (ou DMP, Data Management Plan) **permet de réfléchir à la gestion des données d'un projet en amont afin de l'anticiper** :

- **Quelles données vont être obtenues** : quels types de données, comment sont-elles collectées, où les stocker, comment on sécurise le stockage, quelle volumétrie, quels formats, quelle organisation, quelle documentation ...
- **Comment vont-elles être utilisées** : comment on les partage, comment on les traite, où on les traite, ...
- **Comment elles vont être préservées** : à quel terme, quelles données, où, comment ...
- **Comment elles vont être valorisées** : comment les diffuser, sous quel format, sous quelle licence, quelles données, comment associer les codes, ...
- **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

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

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

Reusable / *Réutilisable*

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



DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

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

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

- Description courte de chaque type de données
 - déjà existant ou non, conditions d'utilisation
 - issue ou liée à un équipement ou logiciel spécifique
 - contraintes à la manipulation ou à la réutilisation
 - brève description du traitement prévu
- Volumétrie estimée : un ordre de grandeur est suffisant
- Formats utilisés : ouverts, libres et multi-plateformes si possible
- Présence de données sensibles ?
 - besoins en sécurité
 - lien avec le Délégué à la Protection des Données (DPD)



2. DOCUMENTATION ET QUALITE DES DONNEES

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

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

- Utilisation d'un standard de métadonnées
- Méthode de documentation des données : fichiers à part, notebook, cahier de laboratoire, documentation incluses avec les données, etc.
 - les personnes faisant les analyses \neq personnes faisant les acquisitions
- Structure d'organisation des fichiers
- Convention de nommage
- Contrôle qualité des données : répétition, calibrage, diagnostique automatique, revue manuelle, etc.
- **Conseil : il n'y a pas de mauvaises réponses, il suffit de décrire ce qui est fait au sein du projet.**