



DATA
TERRA



ODATIS

Campus de la Mer – 14 mai 2024 /

Plan de Gestion des Données

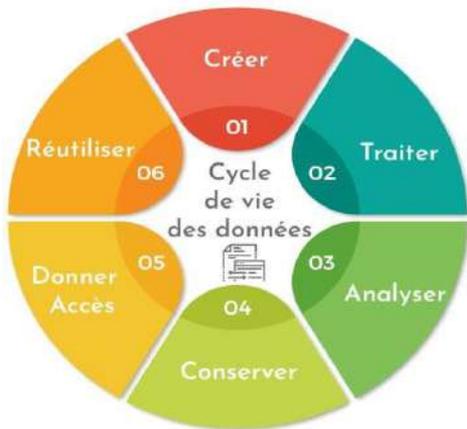
Joël Sudre



Qu'est-ce qu'un PGD

PGD: Plan de Gestion des Données (ou DMP: Data Management Plan)

- Permet de mettre en place les actions nécessaires pour gérer les données d'un projet, d'un service, d'un laboratoire,...
- Permet d'expliquer comment sont gérées les données (création, collecte, partage, archivage)
- Aide à organiser et anticiper toutes les étapes du cycle de vie des données
- Document évolutif : Rédaction dès le début d'un projet (plusieurs versions)



- Acquisition
- Traitement (Calibration)
- Traitement (Validation)
- Distribution
- Pérennisation



Application des principes FAIR

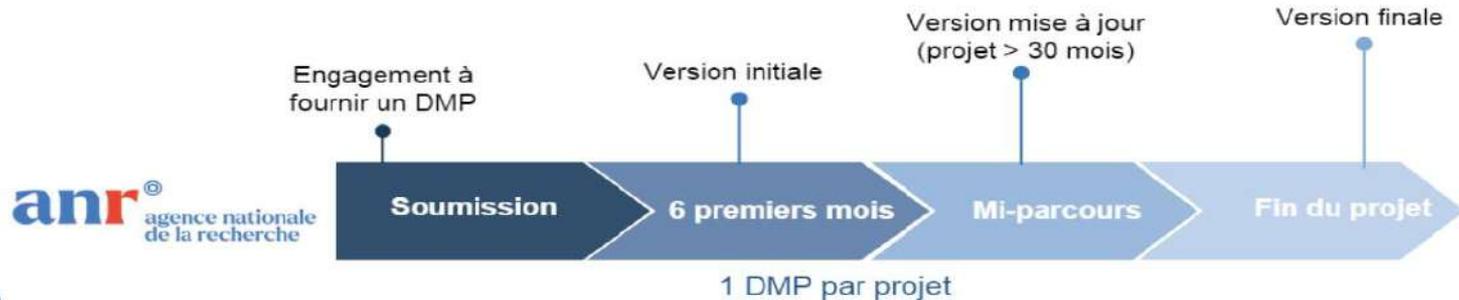
Campus de la Mer, le mardi 14 mai 2024



Intérêts d'écrire un PGD

Pourquoi passer autant de temps sur ce document!?

- Incitation voire Obligation: Organisme de recherche, commission européenne, ERC, Agences de financement, ANR:



- Permet de se poser les bonnes questions pour rendre les données FAIR : Facile à trouver, Accessibles, Interopérables, Réutilisables
- Aide à évaluer les ressources nécessaires
- Définit les responsabilités.

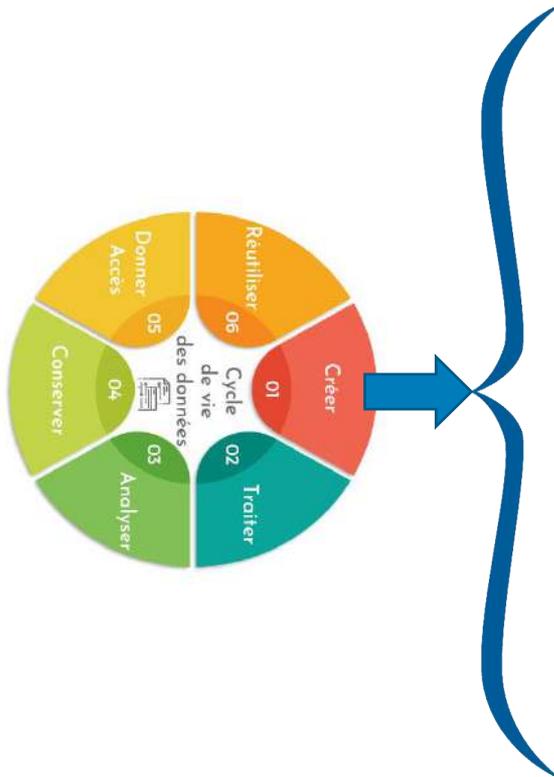
Le Modèle de PGD de l'ANR



1. Documents de référence; 2. Introduction; 3. Contexte
4. **Description des données et collecte ou réutilisation de données existantes**
 - 4.1 Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?
 - 4.2 Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?
5. **Documentation et qualité des données**
 - 5.1 Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?
 - 5.2 Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?
6. **Stockage et sauvegarde pendant le processus de recherche**
 - 6.1 Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?
 - 6.2 Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?
7. **Exigences légales et éthiques, codes de conduite**
 - 7.1 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 ?
 - 7.2 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 ?
 - 7.3 Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?
8. **Partage des données et conservation à long terme**
 - 8.1 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 ?
 - 8.2 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) ?
 - 8.3 Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?
 - 8.4 Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?
9. **Responsabilités et ressources en matière de gestion des données**
 - 9.1 Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?
 - 9.2 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) ?

Campus de la Mer, le mardi 14 mai 2024

Le Modèle de PGD de l'ANR



1. Documents de référence

J'ai créé un nouveau jeu de données

Quels documents de référence je peux fournir:

- Pour que je puisse le reconstruire?
- Pour que je puisse l'utiliser? (description, documentation,...)
- Condition d'utilisation, Limitation d'interprétation des données
- Documents en accès restreint?

2. Introduction

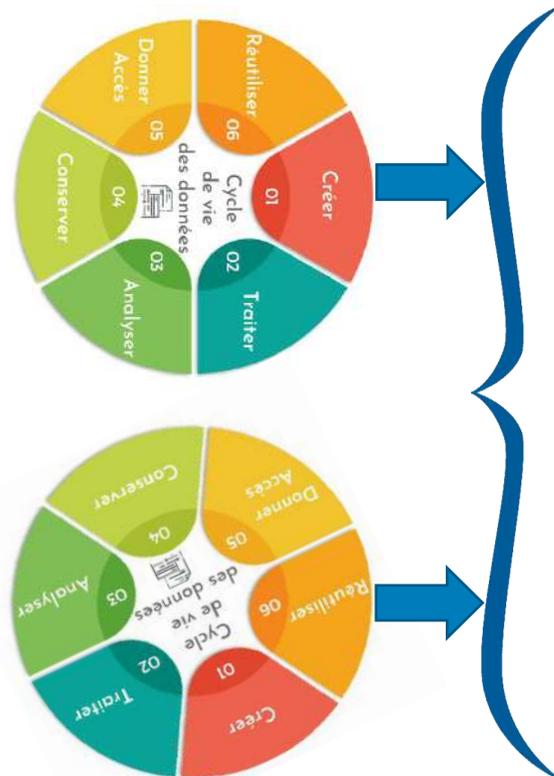
Ce/ces jeu(x) de données sert à quoi? Que mesure-t-il? Description?

3. Contexte

Pourquoi créer ce/ces jeu(x) de données? Que va-t-il apporter?

Campus de la Mer, le mardi 14 mai 2024

Le Modèle de PGD de l'ANR



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

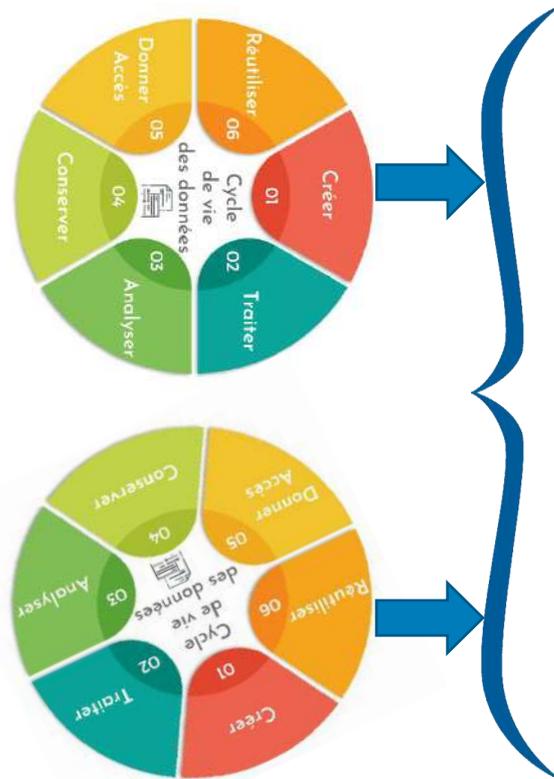
4.1 Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

- Méthodologie / logiciels pour recueillir produire la donnée
- Restrictions à la réutilisation des données préexistantes
- Expliquer la provenance des données

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

- Détails sur le type de données : par exemple numérique (bases de données, tableurs), textuel (documents), image, audio, vidéo, et/ou médias composites
- Détails sur le format (pourquoi ce format), format standard, ouvert
- Les échantillons sont un type de donnée! (support à conserver? Numérisation de l'information)
- Volumétrie (Octets), quantité d'objets, de fichiers, de ligne de colonne etc.

Le Modèle de PGD de l'ANR



5. Documentation et qualité des données

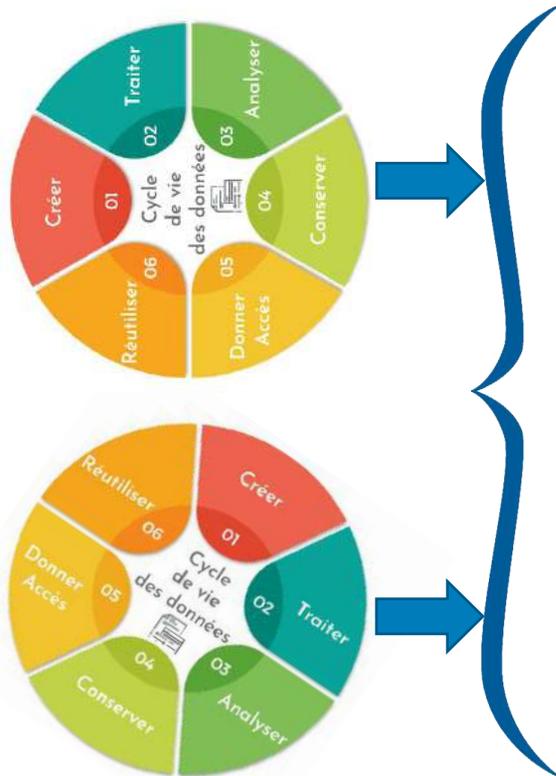
5.1 Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

- Quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données? (standards des métadonnées, convention de nommage, version des données,) - > Voir site <https://www.odatis-ocean.fr/ressources/bonnes-pratiques> (vers la réutilisation)
- Gestion de la donnée: Documentation pour la réutilisation de nouvelles données (méthode de collecte, procédure d'analyse, définition des variables, des unités de mesure, etc.)

5.2 Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

- Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.

Le Modèle de PGD de l'ANR



6. Stockage et sauvegarde pendant le processus de recherche

6.1 Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du **processus de recherche** ?

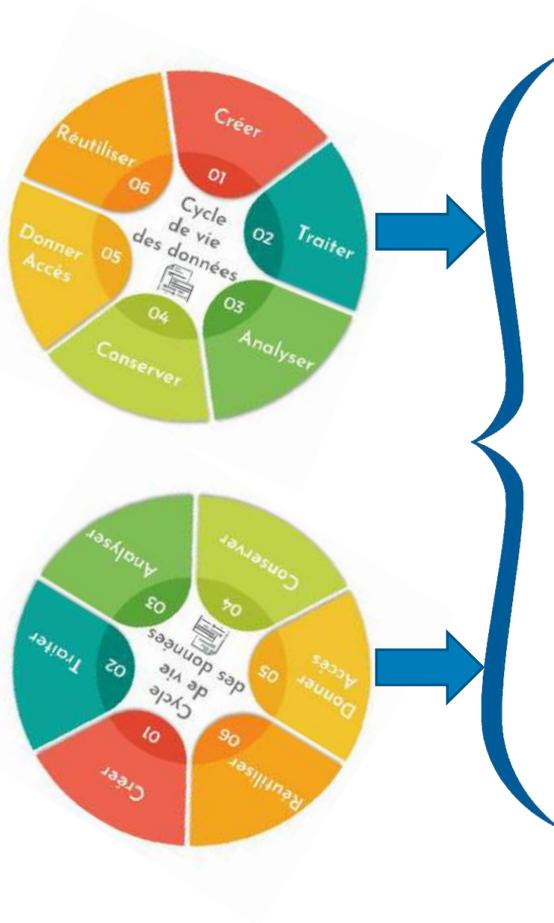
- Type de stockage et de sauvegarde (**règle 3-2-1 trois copies** de vos données; **deux sauvegardes** sur des types de supports différents, et au moins **une sauvegarde** doit être stockée hors site ou sur le cloud), fréquence des sauvegardes
- Disque USB, Disque externe, Clé USB, Disque de portable: **NON!**
- Stockages Robustes
- Campagne à la mer: On sauvegarde sur plusieurs supports, plusieurs personnes ont la sauvegarde, rassemble les données avant de quitter le bord!

6.2 Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du **processus de recherche** ?

- Et en cas de panne? Comment je récupère la donnée?
- Version brute, version traitée, quel version de base je garde? Intérêt ?
- Qui a le droit d'utiliser cette donnée? Contrôle d'accès? Collaboration
- **Attention à la RGPD**
- **Secrets commerciaux**
- **Tenir compte de la politique institutionnelle de la protection des données**

Campus de la Mer, le mardi 14 mai 2024

Le Modèle de PGD de l'ANR



7. Exigences légales et éthiques, codes de conduite

7.1 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 ?

➤ RGPD, pseudonymisation, cryptage, chiffrement

7.2 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 ?

➤ Qui donne accès à la donnée?

➤ Qui en est le propriétaire?

➤ Landing page de DOI, licence (CC), open science ...

➤ Propriété intellectuelle

7.3 Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

➤ Éthique impact sur le stockage de la donnée? Durée de conservation des listes de noms RGPD, qui peut utiliser et/ou voir la donnée?

➤ Comité d'éthique à mettre en place si nécessaire

➤ Code de conduite

Campus de la Mer, le mardi 14 mai 2024

Le Modèle de PGD de l'ANR

8. Partage des données et conservation à long terme

8.1 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 ?

- Dépôt, entrepôt, catalogue -> Pôle ODATIS (plan de préservation)
- Embargo possible: 2 ans à partir du moment que la donnée est scientifiquement exploitable (Pas d'abus SVP!)
- Propriété intellectuelle, dépôt de brevet

8.2 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) ?

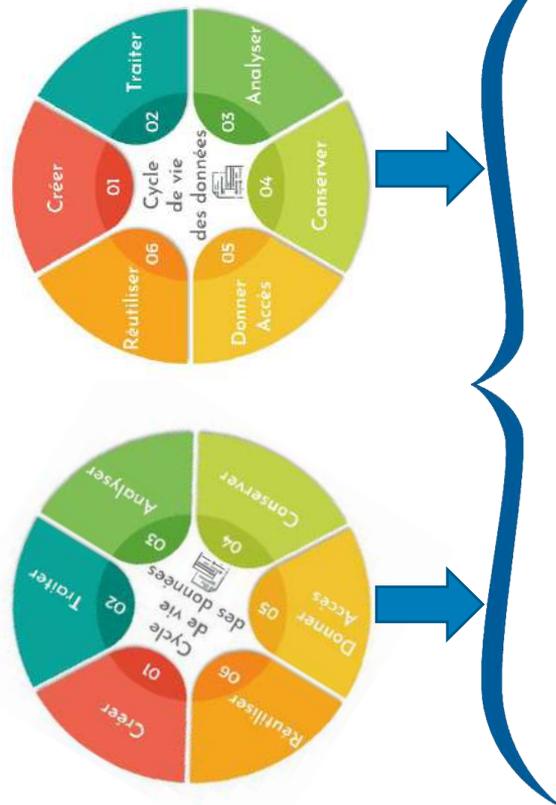
- Donnée sensible divulgable ou non
- Préservée à long terme, détruite (phase de test du capteur)
- Décrire l'utilisation de la donnée et/ou les utilisateurs

8.3 Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

- A discuter avec le pôle ODATIS/SISMER/ARCHIMER/HAL

8.4 Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

- SEANOE/DATAREF (Le DOI devient quasiment OBLIGATOIRE)



Point d'attention sur les Données

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

**STOCKAGE ET
SAUVEGARDE**

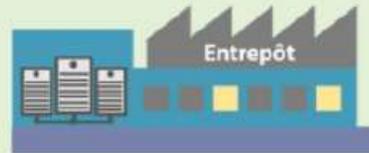


Serveur sécurisé
de l'institution

Espace collaboratif
de travail



PARTAGE



Entrepôt généraliste
ou disciplinaire

ARCHIVAGE PÉRENNE



Plateforme d'archivage
(ex : CINES)

Campus de la Mer, le mardi 14 mai 2024

Quelques informations complémentaires

- ✓ Data Terra : Centre de de référence thématique de Recherche Data Gouv pour les données du Système Terre:

<https://recherche.data.gouv.fr/fr/page/centres-de-reference-thematiques-expertises-par-domaine-scientifique>

- ✓ Recensement des services accompagnant la rédaction des PGD au sein des établissements d'enseignement supérieur et de la recherche:

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

- ✓ Rédaction de votre PGD (fortement conseillé): <https://dmp.opidor.fr/>

Utilisation du MADMP (PDG machine actionnable - <https://dmp.opidor.fr/help#content>)

- ✓ Aide à l'amélioration du contenu de votre PGD : [DoRANum – Grille de relecture de PGD](#)

- ✓ Ressources pédagogiques DoRANum: <https://doranum.fr>

Campus de la Mer, le mardi 14 mai 2024



Merci pour votre attention

