

# Plan de gestion des données : enjeux et méthode

## Contexte et politiques actuelles



## Le PGD/DMP, qu'est-ce-que c'est ?

Plan de Gestion de données (PGD) ou Data Management Plan (DMP)

- Document qui permet de **mettre en place les actions indispensables** pour gérer les données de recherche
- **Décrit la façon dont les données sont gérées** depuis leur création, collecte jusqu'au partage et archivage
- **Evolutif** : la rédaction commence dès le début du projet et peut/doit être mis à jour tout au long du projet
- **Aide à organiser et anticiper** les étapes du cycle de vie des données

**Un outil au service du projet de recherche**

## Un document incontournable



**A l'échelle mondiale :** il est de plus en plus recommandé

**A l'échelle européenne :**

- Exigence de la Commission Européenne



- Outils et infrastructures déployées pour (par exemple OpenAIRE et EOSC)

**A l'échelle nationale :**

- PNSO 1 & 2



- Incité ou obligatoire selon les organismes de recherche, universités, agences de financements

## A quoi sert le PGD/DMP ?

Montrer **comment les données sont gérées** tout au long du projet et au-delà

- Selon de bonnes pratiques
- Dans le respect d'un cadre éthique et juridique

Respecter les **principes FAIR** : « aussi ouvert que possible, aussi fermé que nécessaire »

- Faciliter les échanges entre partenaires
- Anticiper les besoins et gagner du temps

## Ses bénéfices



Réfléchir et évaluer les  
ressources nécessaires



Rendre les données  
*action-machinable*



Définir les rôles et  
responsabilités de chacun



Opportunité de dialogue  
entre les différents  
acteurs



Rendre les données FAIR



Faciles à trouver



**F**



Intéropérables



**I**

**A**



Accessibles



Réutilisables

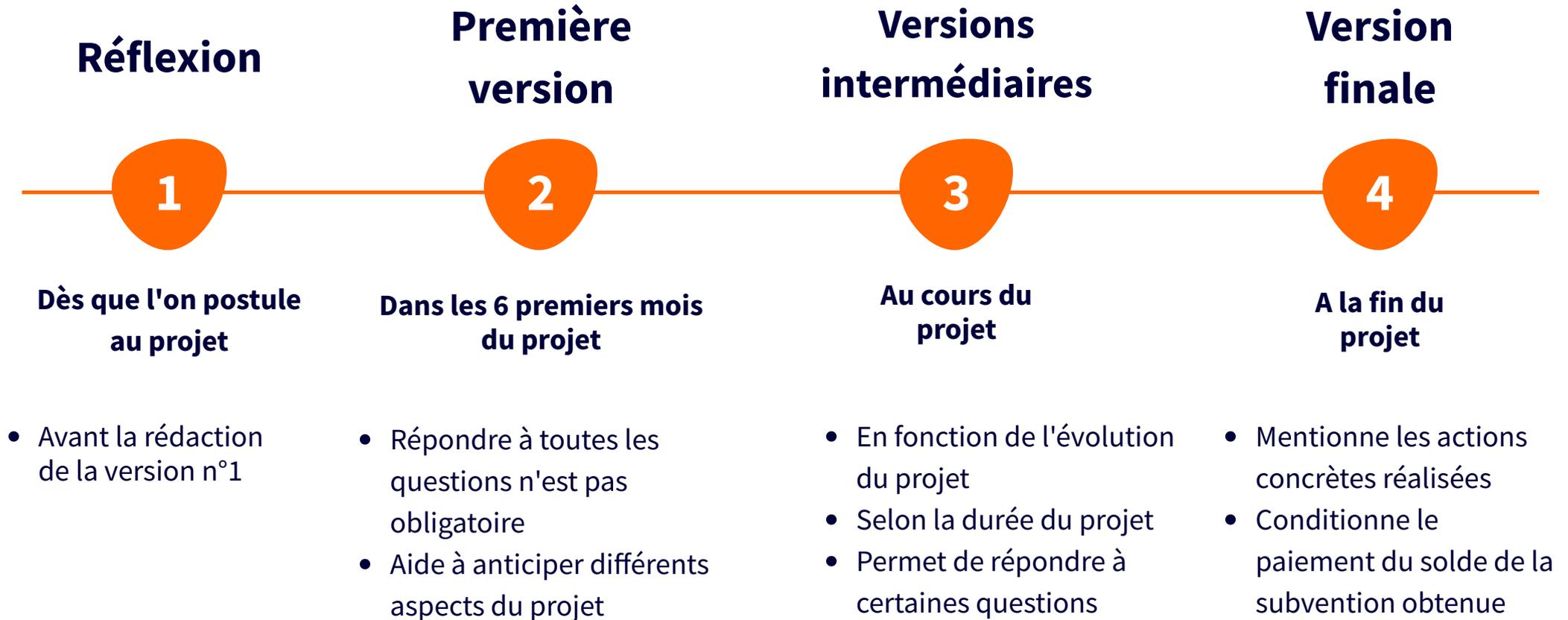


## FAIR data VS Open data

- **Les données ouvertes (open data)** : contenus **librement accessibles** qui peuvent être **utilisés, modifiés et partagés** par n'importe qui pour n'importe quel but
- **Les données FAIR (FAIR data)** : données qui suivent un certain nombre de **principes** pour être **lisibles** et **utilisables** par des **humains** comme par des **machines**

*"Aussi ouvert que possible, aussi fermé que nécessaire"*

## Quand faire un PGD/DMP ?



## Quel contenu dans un PGD/DMP ?

- **Description** des données
- **Documentation** et **qualité** des données
- **Stockage** et **sauvegarde** des données en cours
- Exigences **légales** et **éthiques**
- Stratégie de **partage** et **ouverture** des données
- **Préservation** à long terme
- **Responsabilité** et **ressources**

## Quel contenu dans un PGD/DMP : exemples

### Description des données

« 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 regroupés en dossiers par auteurs, c'est-à-dire 6 dossiers de maximum 30 Mo [...] nous disposons de 119 fichiers .txt (16+103) des lemmatisations [...]

Exemple extrait du DMP public du projet « Hospitam », modèle ANR, de L. Locatelli

## Quel contenu dans un PGD/DMP : exemples

### Documentation et qualité des données

« [...] Nous avons déployé l'**outil de stockage de la TGIR Huma-Num ShareDocs** en formant les membres du projet. [...] Nous avons **1 To de disponible sur cette plateforme**.

L'**organisation par dossier** a été pensé en équipe et prend la forme suivante : **AAP\_ANR ; AXE1** [...].

Cette représentation numérique est **stockée sur ShareDocs**, par terrain et par habitant en utilisant le **codage suivant** :

Ex : **E2GF\_B1= Enquête 2 Grenoble Flaubert Baluchon 1** »

## Quel contenu dans un PGD/DMP : exemples

### Stockage et sauvegarde des données en cours

« [...] Les données biologiques seront stockées dans trois disques durs externes conservés dans des endroits distincts à accès contrôlé et dans un serveur dédié à la sauvegarde [...].

Les données génétiques seront stockées sous forme de fichier .txt [...] .

L'hébergement [...] avec une localisation des données assurées en Union Européenne, à travers un entrepôt de données de type SQL. Une sauvegarde est effectuée toutes les nuits, avec un format agnostique du moteur de base de données (JSON). »

## Quel contenu dans un PGD/DMP : exemples

### Exigences légales et éthiques

« Le fichier informatique utilisé [...] est mis en œuvre conformément à la **règlementation française (loi Informatique et Libertés modifiée) et européenne (Règlement Général sur la Protection des Données–RGPD)**. [...] aucune Investigation Clinique ne peut être pratiquée sur une personne **sans son consentement libre et éclairé, recueilli par écrit** après que lui a été délivrée l'information prévue à l'article 63, paragraphe 2 du Règlement Européen 2017/745. »

## Quel contenu dans un PGD/DMP : exemples

### Stratégie de partage et ouverture des données

« Comme il s'agit de données personnelles et potentiellement sensibles, un partage sur un entrepôt de données n'est pas envisagé de façon ouverte. Les données récoltées seront conservées en France pendant 3 ans après la fin du projet et anonymisées, [...]. Seuls les membres du projet scientifique ANR PROSECO auront accès aux baluchons et à leur version numérisée de manière sécurisée. »

## Quel contenu dans un PGD/DMP : exemples

### Préservation à long terme

« Les données cliniques seront **conservées en base active jusqu'à 2 ans** après la dernière publication de la recherche. Les données biologiques **seront conservées sur le CLE jusqu'à 10 ans après** la dernière publication de recherche. Les métadonnées [...] être **archivées sur le site d'archivage EGA**. Un DAC contrôlera l'accès aux données anonymisées à des tiers pour des recherches scientifiques dans l'intérêt public. »

## Quel contenu dans un PGD/DMP : exemples

### Responsabilités et ressources

« Le responsable de la gestion des données est le Chef de projet RHU PrediMAP- APHP. Les personnes impliquées dans le traitement des données cliniques :

- Technicien de recherche clinique pour la saisie des données clinique
- Attaché de Recherche Clinique : gestion et traitement des queries
- Datamanagement programmation de la base de données / eCRF

[...]

- 120 000 euros pour la supervision des aspects techniques de la gestion des données de la recherche clinique [...] »

## Les outils pour faire un plan de gestion de données



### DMP Tool

- Créer par l'Université de Californie
- Modèles simplifiés de DMP



### DMP Online

- Créé par le Digital Curation Center (DCC)
- Prend en compte les recommandations de la Commission Européenne (H2020)



### DMP Opidor

- Instance française de DMP Online du DCC qui est mis à disposition de l'ESR
- Hébergé et géré par l'Inist-CNRS



- Gratuit
- Accessible à l'ensemble de l'ESR et partenaires français ou étrangers
- Sécurisé
- Evolutif

- **Modèles financeurs et d'établissements** disponibles
- **Référentiels pour certains modèles**, par ex Science Europe, qui permet de fournir des listes standardisés de termes
- Fichier **téléchargeable sous différents formats**
- Espace institutionnel AMU
- **Assistance personnalisé de rédaction**



## Pour conclure ... quelques petites astuces pour rédiger

1

Commencer tôt

2

Réfléchir à la réutilisation des données

3

Vérifier les politiques en vigueur

4

Se faire aider par les services d'appui

5

Penser large

6

S'inspirer du travail d'autrui

7

Etre précis là où c'est nécessaire

8

Etre concret

9

Oser avouer qu'on ne sait pas  
(encore)

10

Mettre à jour

# Merci pour votre attention !

## Des questions ?

Assistance à la rédaction et relecture de PGD/DMP :  
[scd-dmp-assistance@univ-amu.fr](mailto:scd-dmp-assistance@univ-amu.fr)