

# Les plans de gestion de données Entité

**Francis Ogereau :**

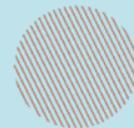
Ingénieur Mésocentre / Cellule Science Ouverte

13 Juin 2024

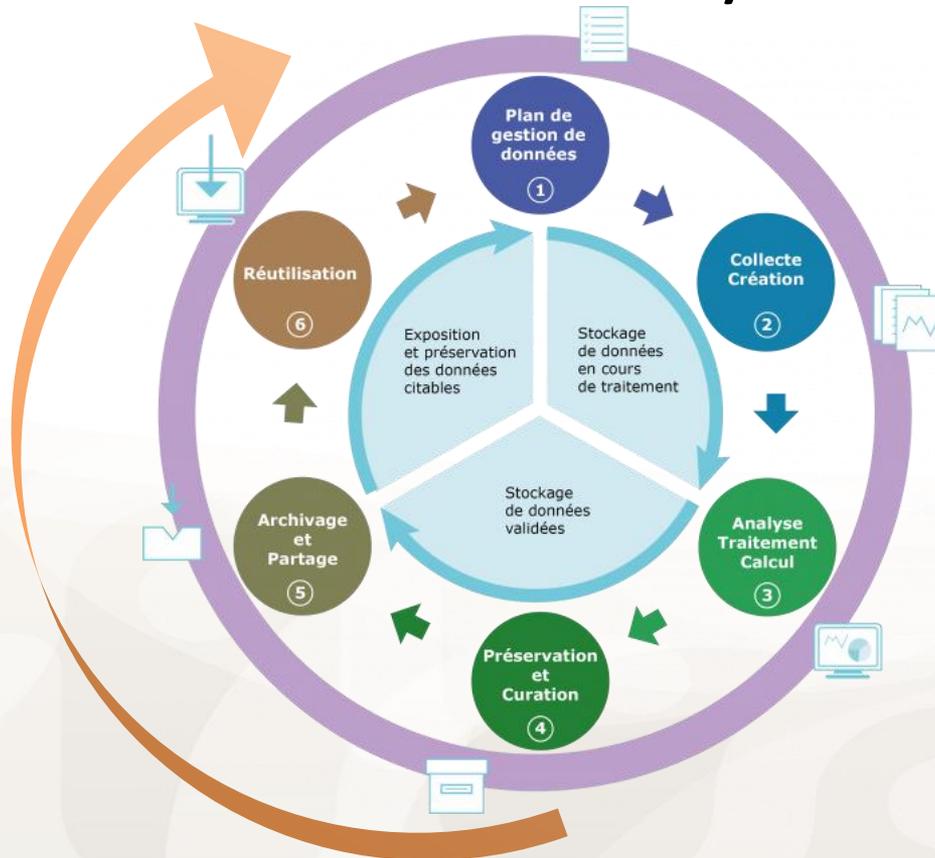




# Cycle de vie de la donnée



# Cycle de vie

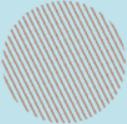


- [doranum-inrae](https://doranum-inrae)
- [harvard](https://harvard)
- Check list : <https://osf.io/d8fqh>



# Plan de gestion de données



- 
- 
- Données de la Recherche
  - PGD
  - PGD d'entité



# Les données de la Recherche

« Les données de la recherche, ou données de recherche, ou données scientifiques, peuvent être définies comme l'ensemble des informations collectées, observées ou créées sous une forme numérique dans le cadre d'un projet de recherche »

- Données d'observations
- Modèles et simulations
- Données dérivées ou compilées
- Code informatique
- Données expérimentales

# Les Plans de Gestion de données

C'est un document synthétique qui aide à organiser et anticiper toutes les étapes de vie du cycle de vie de la donnée. Il explique pour chaque jeu de données comment seront gérées les données d'un projet, depuis leur création jusqu'à leur partage et leur archivage.

# Les Plans de Gestion de données

- Il n'existe pas de trame unique
- Outil d'aide à la rédaction comme DMP Opidor ou Dsw
- Le degré de précision ou l'agencement peut varier d'un modèle à un autre.

## PGD

Plan de gestion de données



# Les Plans de Gestion de données

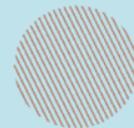
- Les PGD d'entité
  - Décrit les données créées/observées/collectées
  - Définit comment sont gérées les données au sein d'une entité de recherche
  - Bonnes pratiques de gestion/ouverture des données
  - Contribue à la **FAIRisation** des données
  - Simplifie la rédaction d'un PGD de projet
  - Est un document évolutif



# DMP Opidor



*Site web pour créer des PGD*



# DMP OPIDOR

- Site web permettant la création de PGD
- V4 le 4 avril incluant les PGD d'entité
- Propose des modèles et des exemples

# DMP OPIDOR

- **Présentation de l'outil**
- <https://dmp.opidor.fr/>
- <https://dmpopidor-preprod.inis.fr/>

**Merci de votre attention !**



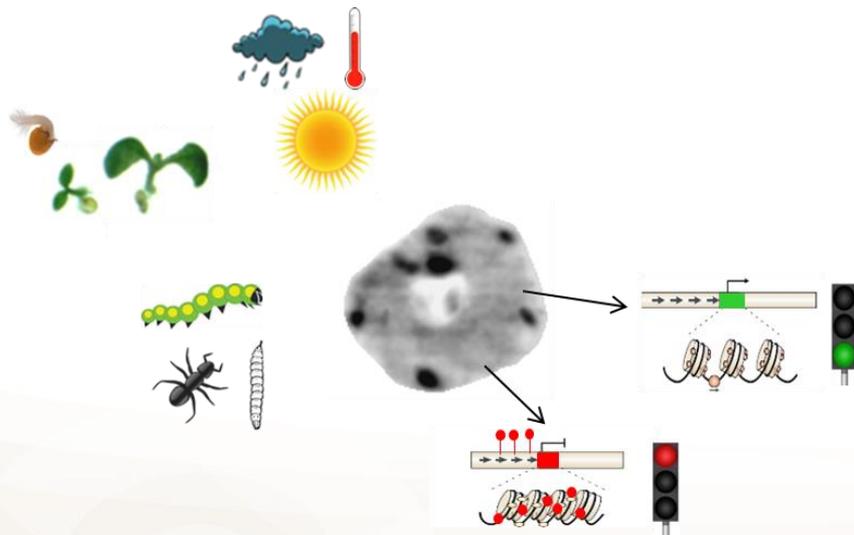
**CLIC**  
Clermont-Ferrand Imagerie Confocale  
Microscopie – Imagerie

## Retour d'expérience Du Plan de Gestion de Données entité

Sophie DESSET



# CODED

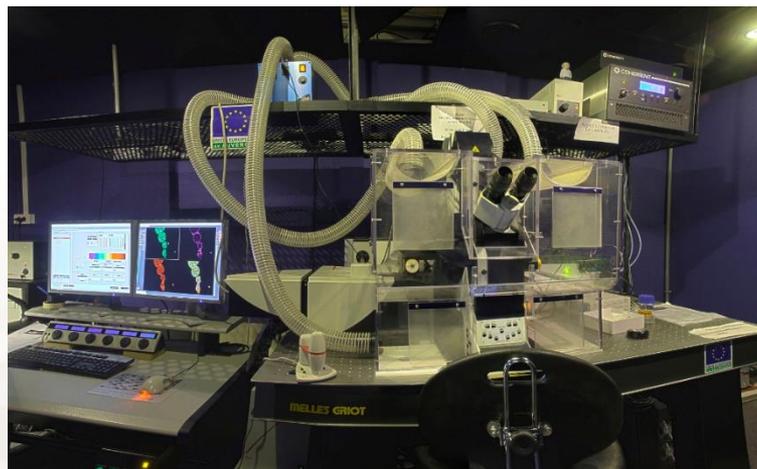
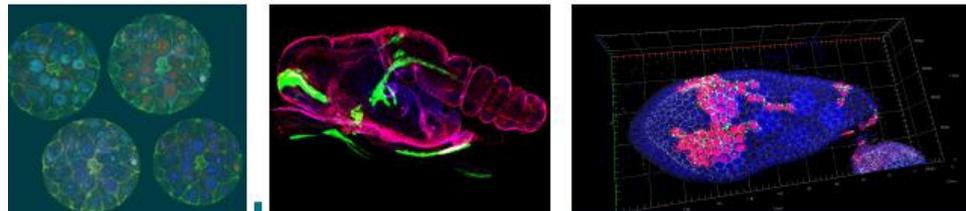


*Arabidopsis thaliana*

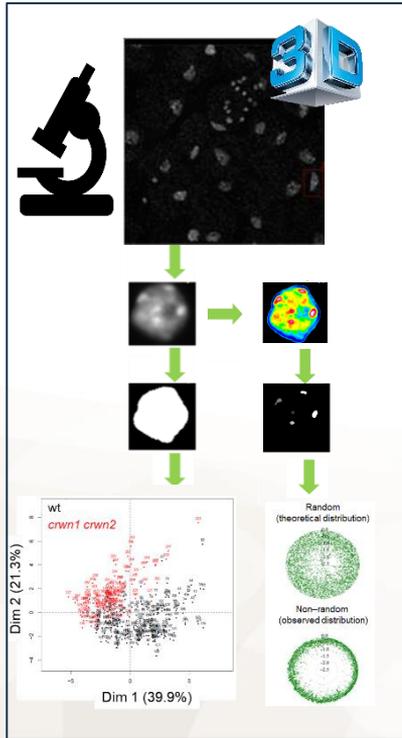


# Clermont Confocal Imaging (CLIC)

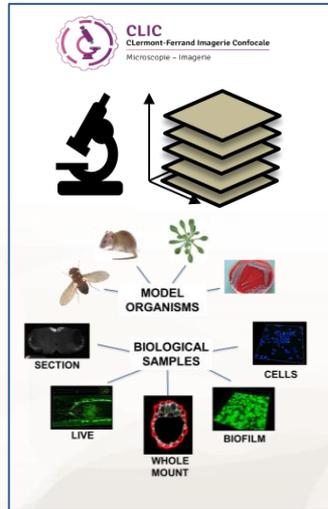
...



# BESOINS 2012-2017



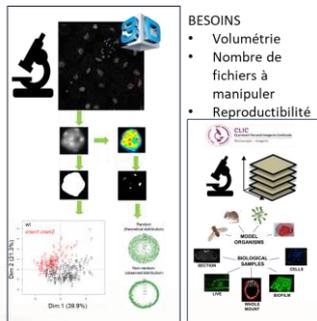
- Ressources informatiques
- Visualisation, tri, nomenclature,...
- Reproductibilité



BESOINS  
2012-2017

PROJETS 2017-2023

VERROUS



Projet de structuration  
NOYAU-HD



Réseau noyaux de plantes



Projet doctoral Intelligence  
artificielle



OXFORD  
BROOKES  
UNIVERSITY



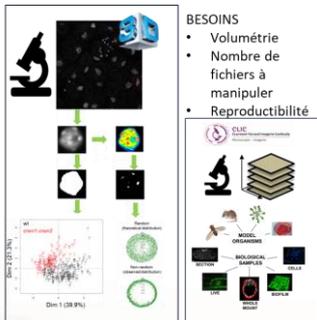
UNION EUROPÉENNE  
Fonds européen de  
développement régional

- Sécurité: Dupliquer les données
- Publication des données : pas d'entrepôt, pas de norme pour les métadonnées
- Codes non reproductibles: ex 6 méthodes DL partageables/106 publications
- Absence jeux de données labellisés publics
- Pas d'anticipation sur les ressources dans les projets

BESOINS  
2012-2017

PROJETS  
2017-2023

En 2024



- Plan national Science Ouverte 2021-2024



- Collaboration avec AuBi



- Gestionnaire d'images au mésocentre

- Entrepôt d'images en Floride

L'implication en science ouverte répond à  
une nécessité scientifique

## Comment rendre les images FAIR ?

- Quelles métadonnées, sous quelle forme?
- **Plan de gestion de données (PGD)?**
- Quel entrepôt pour la publication?
- Quel coût pour le stockage, l'archivage,...?



Projet européen Elixir 2021-2028

# MUDIS4LS

Mutualised Digital Space for FAIR Data in Life and Health Science

*Institut Français Bio-informatique*

•IS 1. Imaging data integration and FAIR sharing.



Pierre  
Pouchin



Nadia  
Goué



Christophe  
Tatout

## Définition Plan de Gestion de données (rappel)

- Le PGD est un document synthétique qui aide à **organiser et anticiper** toutes les étapes de vie du cycle de vie de la donnée. Il explique pour chaque jeu de données comment seront gérées les données d'un projet, depuis leur création jusqu'à leur partage et leur archivage.

## Plan de Gestion de données ENTITE (structure)

- **Entité**: PGD pour un laboratoire, une plateforme, une unité de service  $\neq$  PGD projet.
- Le PGD entité est une aide / un **préalable à la rédaction des PGD projet**.
- Les PGD sont toujours **EVOLUTIFS**

## Formulaires de Plan de Gestion de données

- Il n'existe pas de trame unique.
- Le degré de précision ou l'agencement peut varier d'un modèle à un autre.
- Deux outils de saisie centralisée des PGD:
  - En 2022, **DSW** (Data StewardShip Wizard) était un outil où on pouvait modifier le formulaire, mieux adapté pour les entités.
  - **Opidor** est issu des sciences sociales, il était davantage dessiné pour les projets.
  - **MUDIS4LS a choisi DSW.**
- Dès le départ était posé le problème d'interopérabilité entre les 2 outils.
- **Investissement local au sein de MUDIS4LS pour concevoir la trame bio-imagerie avec plusieurs autres plateformes via la rédaction du PGD Au-Bi.**

## Comment rédiger un PGD structure ou entité?

- Au commencement... il n'y avait rien!

-> La première étape a consisté à rédiger le FORMULAIRE du PGD entité sur DSW

# Réalisation et étude d'une triple cartographie

## ❑ **Cartographie applicative :**

- Connaître la nature des données collectées au sein de l'entité + leur niveau de sensibilité

## ❑ **Cartographie informatique :**

- Moyens techniques et informatiques dont les agents disposent au sein de l'entité
- Etat des lieux des pratiques de sauvegarde/stockage des données auprès d'un échantillon représentatif (corps, institut...) des utilisateurs de l'entité.

## ❑ **Cartographie institutionnelle :**

- Qui fait quoi
- responsabilités

## Projets

Rechercher les projets... Date de dernière modification ↑ Modèles de projets Étiquettes du projet Modèles de connaissances **Créer**

sophie desset ▾

- A** **AuBi Bioimagerie** 👤👤👤👤 +2 · PGD structure développé par l'IFB 2.6.3
- C** **CLIC - UCA** 🧠🧠🧠🧠 **mise à jour disponible** · PGD structure développé par l'IFB 2.4.25 *Mise à jour il y a environ 2 mois* ⋮

2 PGD entité / structure bioimagerie rédigés avec Nadia Goué, Pierre Pouchin, Francis Ogereau

- a. Bioimagerie**
- b. Cytométrie**
- c. Génomique**
- d. Métabolomique**
- e. Protéomique**

Les trames d'autres thématiques ont été rédigées dans le cadre du MUDIS4LS



DSW@IFB

**AuBi Bioimagerie**   

Questionnaire Métriques Aperçu Documents Paramètres

Lire Importation de réponses

À faire 5 Commentaires Historique des versions 

Tableau de bord

Modèles de connaissances

Modèles de documents

**Projets**

Liste

Actions

Importateurs

(FR) Mentions légales

(FR) Données personnelles

(EN) Legal notice

(EN) Personal data

**Phase en cours**

Avant soumission du PGD

**Chapitres**

I. Introduction 

► Pour chaque domaine disciplinair...

II. Informations générales 

III. Données de la recherche **16**

# I. Introduction

Les objectifs de ce plan de gestion des données (PGD):

- Accompagner les plateformes et leurs utilisateurs dans le cycle de vie de leurs données ;
- PGD structure et PGD projet : nous proposons ici les champs à renseigner pour le PGD structure, accompagnés de recommandations ; ce PGD structure a pour but de venir renseigner (automatiquement, voir le prochain point) certains éléments des futurs PGDs projet des utilisateurs de la structure ;
- Machine actionnable data management plan (maDMP) : ce PGD est proposé dans un format qui se veut compatible avec le projet maDMP de l'INIST et de l'IFB, qui permettra de réutiliser et partager facilement les briques de description (par exemple pour la mise en place de l'espace de stockage d'un projet, pour nourrir le PGD d'un projet...), afin d'éviter la double saisie et s'assurer que les éléments pertinents sont connus des acteurs concernés.

Dans ce contexte, la structure est définie d'une part comme l'infrastructure qui est génératrice de données à la demande des projets de recherche, et d'autre part, comme l'entité qui est la garante de la pérennité des données une fois le projet arrivé à son terme.

Le PGD structure a pour vocation de servir ensuite de modèle à tout PGD de projet porté par l'infrastructure.

Le document qui suit décrit les éléments du cycle de vie de la donnée numérique, c'est-à-dire, les processus de création, d'identification, de documentation, de partage et d'archivage. Par souci de simplification pour cette première version du PGD structure, certaines caractéristiques constitutives sont imposées et certains types de données en sont exclus à ce stade. Néanmoins ce document, évolutif et destiné à être modulaire, s'enrichira en fonction des exigences et recommandations de diverses entités tutélaires, et des besoins émanant des utilisateurs actuels et potentiels.

L'objectif premier du PGD structure est d'accompagner les données de tous renseignements permettant ainsi de les rendre ainsi visibles, accessibles, et réutilisables quel que soit le demandeur. Cet objectif s'inscrit dans le paradigme FAIR (<https://www.go-fair.org/fair-principles/>), soit « Facile à découvrir, Accessible, Interopérable, Réutilisable », qui est l'un des piliers de la science ouverte.

Un autre objectif est l'automatisation de la collecte et de la diffusion des informations pertinentes aux données, afin d'optimiser ce processus fastidieux, de le rendre plus robuste en limitant l'intervention humaine, et, enfin, de garantir une visibilité maximale par l'exposition du cadre qui gouverne la réutilisation des données.

Ce projet « PGD structure » est une extension du projet « PGD structure bioimagerie », porté par et réalisé conjointement avec les Infrastructure Nationale en Biologie Santé (INBS) « France-Bioimaging » (FBI), « Institut Français de Bioinformatique » (IFB), et « Centre National De Ressources Biologiques Marines » (EMBRC-France).

## PERPECTIVES

- Traduction DSW bio-imagerie en anglais en cours (MICA-CLIC).
- Développement Opidor: trames PGD entités disponibles (mais non spécifiques).
- Collaborations Opidor avec des porteurs de PGD structure.
- Interopérabilité Opidor/DSW a commencé et soutenue par l'IFB officiellement.

## RETOUR EXPERIENCE SUR LES PGD

- Multidisciplinaire, **à faire à plusieurs**.
- Permet une vraie réflexion sur les pratiques et les besoins futurs des entités.
- Réelle assistance à l'anticipation pour la réalisation des projets.
- Usage des PGD entité et rédaction des PGD projets malheureusement vécu comme une contrainte par la communauté qui ne s'en saisit pas

Et maintenant?

- Finir PGD CLIC -> c'est pour nous
- PGD laboratoires? Autres plateformes? -> **c'est à vous!**

**Merci de votre attention !**

# Les données de la recherche

- les **données d'observation** comme des relevés de température ou de précipitations, les données de télédétection, les photos d'un événement ou encore les données d'enquêtes. Elles sont uniques et impossibles à reproduire d'où la nécessité de les conserver indéfiniment.
- les **données expérimentales** comme celles générées par un équipement en laboratoire ou les mesures de performance d'une machine. Ces données sont supposées être reproductibles mais il peut être difficile de reproduire les mêmes conditions de l'expérience ou reproduire la manipulation peut être tellement coûteux que cela en devient prohibitif. Par conséquent, il est recommandé d'archiver avec soin les données recueillies dans de telles situations.
- les **données de simulation numérique** produites par des logiciels comme pour des modèles climatiques ou économiques. Ces données sont reproductibles et s'il n'est pas nécessaire d'archiver toutes les données ainsi produites, il est en revanche indispensable de décrire le modèle, le logiciel associé et d'en archiver toutes les informations qui permettront sa ré-exploitation.
- les **données dérivées ou compilées** comme celles obtenues par de la fouille de textes ou de données d'une base. Il s'agit de données brutes qui ont fait l'objet de traitements et d'analyses successifs. Il est fortement recommandé de les archiver avec soin.
- le **code informatique** : les enjeux liés à l'Open Data, aux infrastructures de diffusion ou au contexte juridique le rapprochent des données.



**FAIR**



Let's GO FAIR

<https://www.go-fair.org/>

RDA

<https://www.rd-alliance.org/>



# Principes FAIR

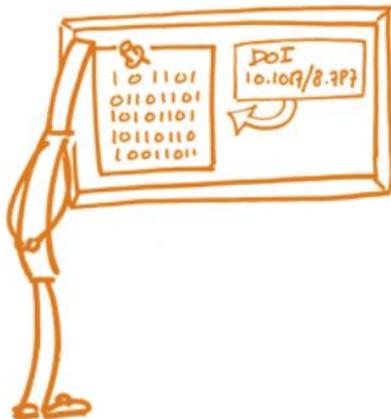
Crédits : open-science-training-handbook.github.io

## FAIR DATA PRINCIPLES

AH!



FINDABLE



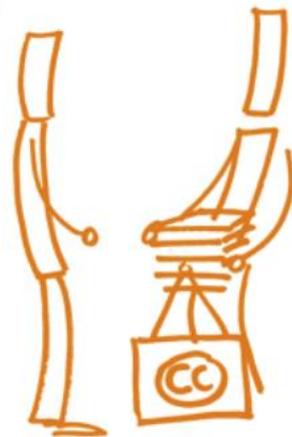
ACCESIBLE

HOW DO YOU  
OPEN A .XEQ FILE?



INTEROPERABLE

- HERE



REUSABLE

# Principes FAIR

