



**HAL**  
open science

## Les Plans de Gestion des Données entités

Jean-François Dufayard, Paulette Lieby

► **To cite this version:**

Jean-François Dufayard, Paulette Lieby. Les Plans de Gestion des Données entités. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParis-Tech; Université de Lille; Sorbonne Université; Data Terra, Jun 2024, Paris, France. 10.5281/zenodo.12205835 . hal-04683415

**HAL Id: hal-04683415**

**<https://hal.inrae.fr/hal-04683415v1>**

Submitted on 2 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Les Plans de Gestion des Données *entités*

Contribution aux supports: Paulette Lieby (IR) - 

Jean-François Dufayard (chercheur) - 



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

*Infrastructure nationale en biologie et santé  
Centre de référence thématique  
pour Recherche Data Gouv*

# De quoi parle-t-on ? Un peu de sémantique



# Un PGD, en bref

PGD :

- Plan de Gestion des Données
- Pour Générer du Dialogue © Frédéric de Lamotte

Définir une **stratégie concertée** de gestion des données, le long de leur **cycle de vie**, dans le cadre d'un **projet** ou d'une **entité pérenne** gérant des données.

# Ce que n'est pas un PGD

**Macro** gouvernance

→ Les lois nationales, internationales

# Ce que n'est pas un PGD

## **Macro** gouvernance

→ Les lois nationales, internationales

## **Micro** gouvernance

→ Politiques de données (des tutelles, des communautés identifiées)

# Ce que n'est pas un PGD

## **Macro** gouvernance

→ Les lois nationales, internationales

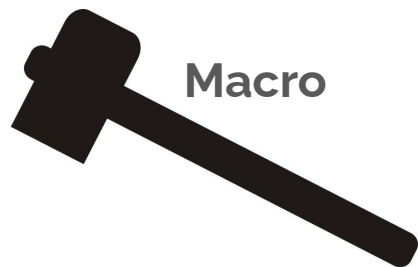


## **Micro** gouvernance

→ Politiques de données (des tutelles, des communautés identifiées)



# Ce qu'est un PGD et sa situation



**Méso** gouvernance

→ Le plan de gestion: notre degré de liberté





# Ce qu'est un PGD et sa situation



**Attention: parfois on veut forger l'enclume**

**À la lumière des PGD entités, on veut éclairer la micro gouvernance**

# PGD projet ? PGD structure ? PGD entité ????

**PGD projet** : une stratégie de gestion, sur la durée d'un projet, mais traitant aussi des données froides, voir de l'archivage... un plan de gestion à **durée déterminée**

**PGD entité** (anciennement structure) : une stratégie de gestion, sur une entité pérenne (plateforme, laboratoire...), un plan de gestion à **durée indéterminée**

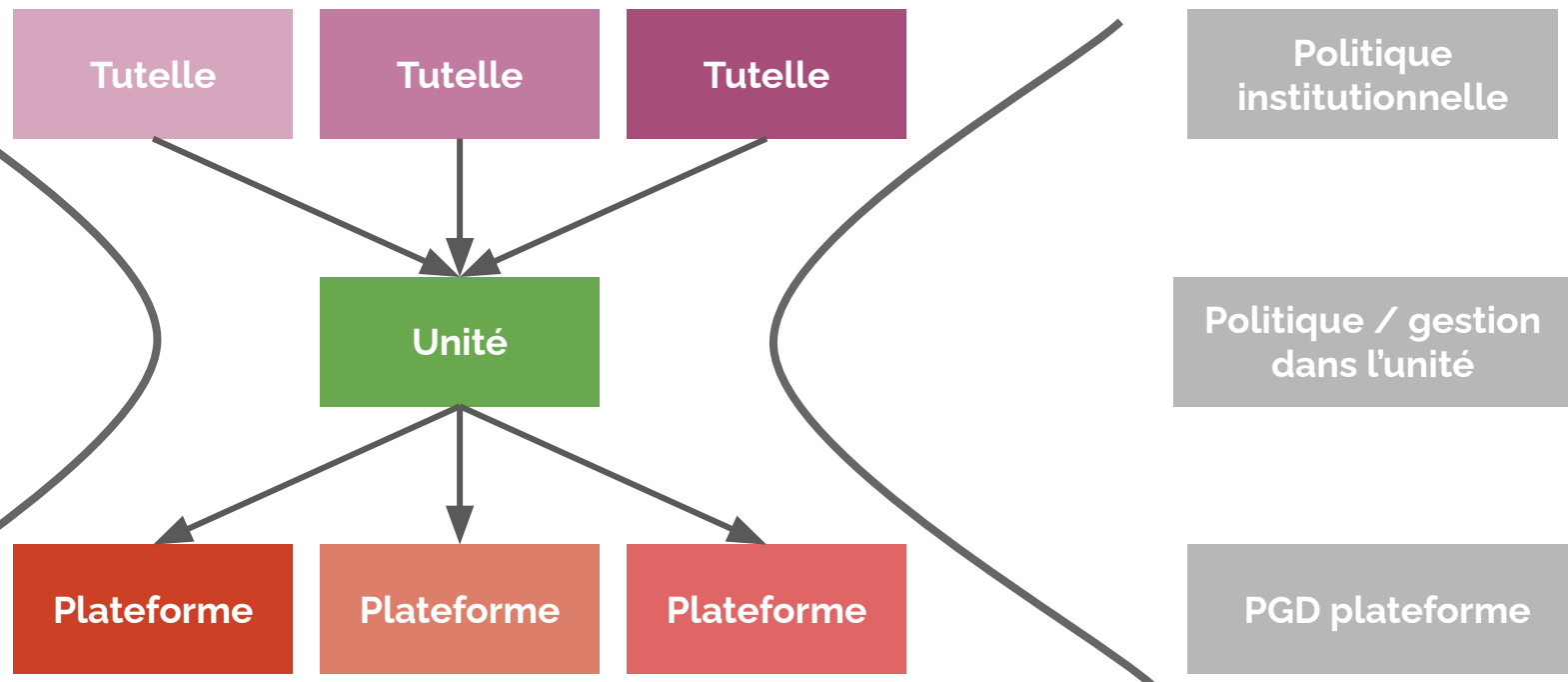
# PGD entité ? Mais pour quelle entité ?

## Entité ?

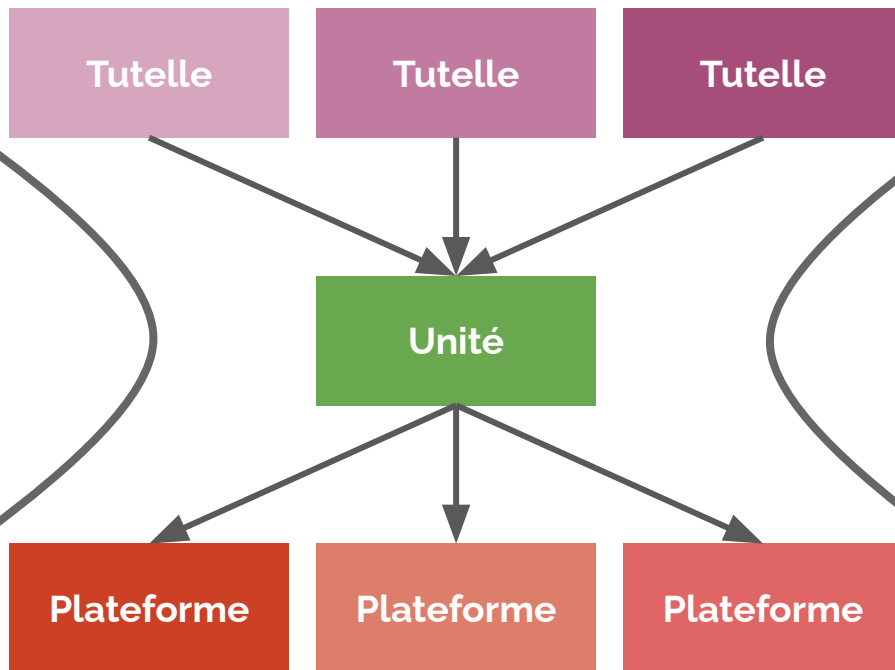
- Plateforme ?
- Plateau ?
- Laboratoire / unité ?
- Institut ???

**Comment et pourquoi aborder l'exercice du PGD pour des entités aussi hétérogènes ?**

# La question ouverte du "sablier"



# La question ouverte du “sablier”



Du document de **politique**, au **PGD**, en passant pas les **recommandations** pour les PGD

# PGD classique, PGD structuré



**Attention au gag de vocabulaire**

**Structure (entité) ≠ Structuré**

**PGD classique** : ne contient que des réponses en texte libre

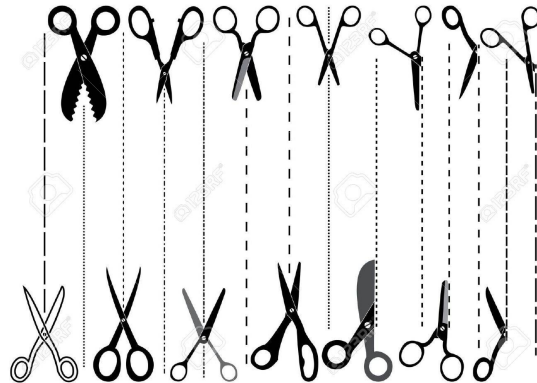
**PGD structuré** : une partie des réponses sont interprétables par une machine (chiffres, dates, choix multiples...)

Exemple: quel est la taille attendu d'un jeu de données ?

Simple: *"oh ben ça dépend, sur les plantes ça peut monter à 40Go, mais sur des petits organismes on va pas dépasser les 10Go... Après on restera loin du To bien entendu".*

Structuré: **25Go**

Le premier pas est le plus difficile:  
les produits de la recherche



# Le nœud gordien : le produit de la recherche

**Produit de la recherche:** ensemble de données gérées de la même façon, dans le cadre d'un projet ou d'une entité

On **découpe** notre PGD en produits de la recherche

On répond aux **questions du PGD** pour chaque produit de la recherche

Il n'existe **jamais une seule solution** de découpage



# Le produit de la recherche - exemple projet

Classiquement, dans un projet, avec de multiples types de données

Par exemple, dans un même projet:

- Des données de bioimagerie
- Des données de séquençage

Produit de la recherche 1

## Bioimagerie

Images brutes - images analysées - images annotées...

Produit de la recherche 2

## Séquençage

Lectures brutes - assemblages - *mappings* - annotations...

...

...

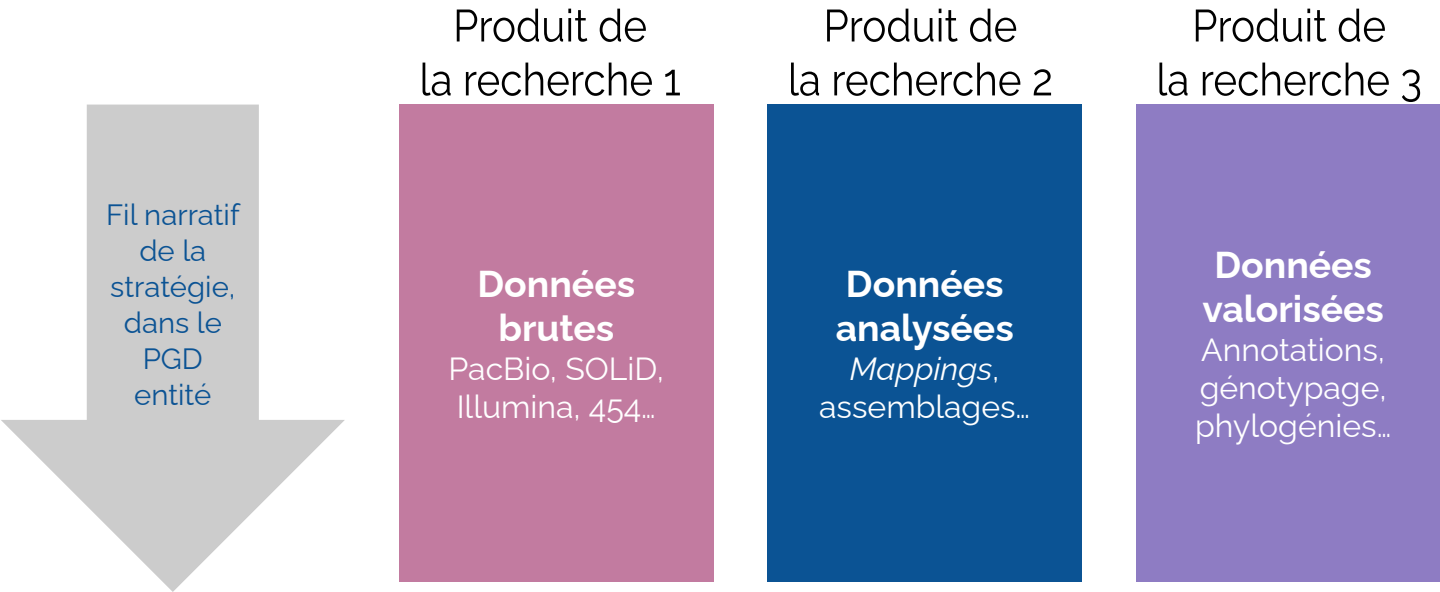
Fil narratif de la stratégie, dans le PGD projet

# Le produit de la recherche - exemple entité

Dans le cas d'une plateforme de production et d'analyse de données

Par exemple, pour une plateforme de séquençage:

- Des données brutes
- Des données analysées (*mappings*, assemblages)
- Des données valorisées (génomomes annotés, génotypage, phylogénies...)

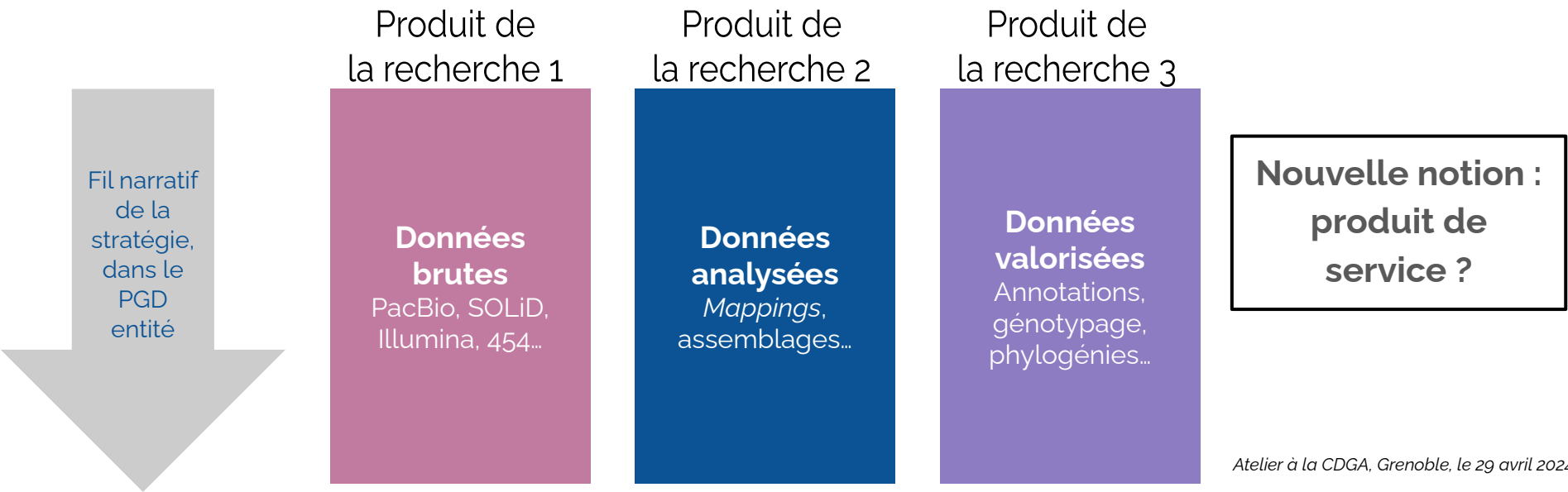


# Le produit de la recherche - exemple entité

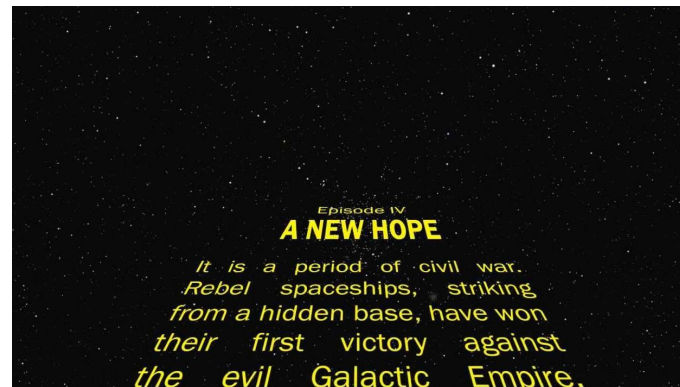
Dans le cas d'une plateforme de production et d'analyse de données

Par exemple, pour une plateforme de séquençage:

- Des données brutes
- Des données analysées (*mappings*, assemblages)
- Des données valorisées (génomomes annotés, génotypage, phylogénies...)



Et c'est grave ?  
Un peu, si on raisonne sur le long terme  
Parlons d'avenir



# Les PGD entités et les PGD projets en synergie

Ce qu'on aimerait dans l'avenir, des PGD opérationnels,  
dits ***“machine actionable”***

PGD entité

PGD projet

# Les PGD entités et les PGD projets en synergie

Ce qu'on aimerait dans l'avenir, des PGD opérationnels, dits ***“machine actionable”***

Quels sont les prochains projets à émarger ?

Quelle évolution des besoins pour mon collectif ?

Combien de nouveaux utilisateurs, quels groupes de travail à administrer ?

Quel est le futur plan de charge du collectif ?

**PGD entité**

**PGD projet**

Quel matériel pour produire les données ?

Quel matériel pour stocker et analyser les données ?

Quels tarifs ?

Comment renseigner mon PGD projet avec l'ensemble de ces informations ?

# Les PGD entités et les PGD projets en synergie

Ce qu'on aimerait dans l'avenir, des PGD opérationnels, dits ***“machine actionable”***

Quels sont les prochains projets à émarger ?

Quelle évolution des besoins pour la plateforme ?

Combien de nouveaux utilisateurs, quels groupes de travail à administrer ?

Quel est le futur plan de charge de la plateforme ?

**PGD entité**

**Spécifications techniques  
(matériel de production,  
stockage,  
sécurité...)  
Modèle économique**

...

**Besoins en stockage  
Besoin en calcul  
Durée du projet  
Comptes utilisateurs**

...

**PGD projet**

Quel matériel pour produire les données ?

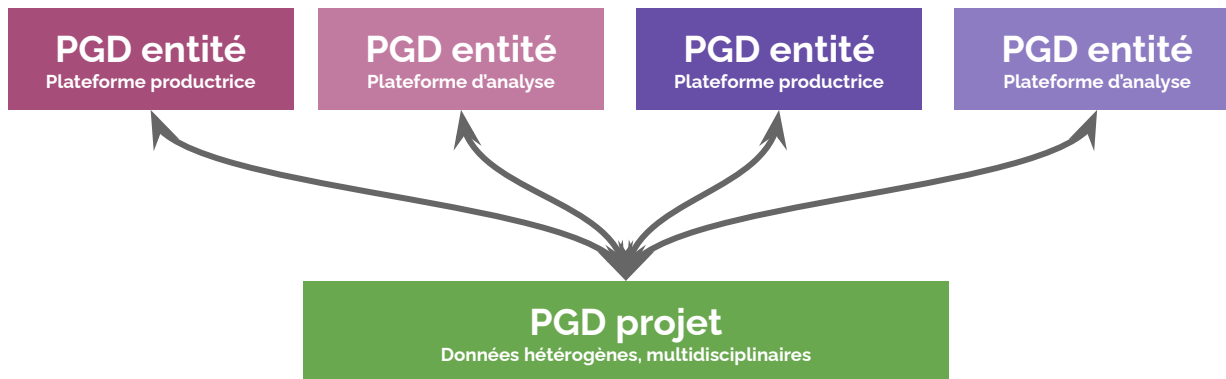
Quel matériel pour stocker et analyser les données ?

Quels tarifs ?

Comment renseigner mon PGD projet avec l'ensemble de ces informations ?

# Les PGD entités et les PGD projets en synergie

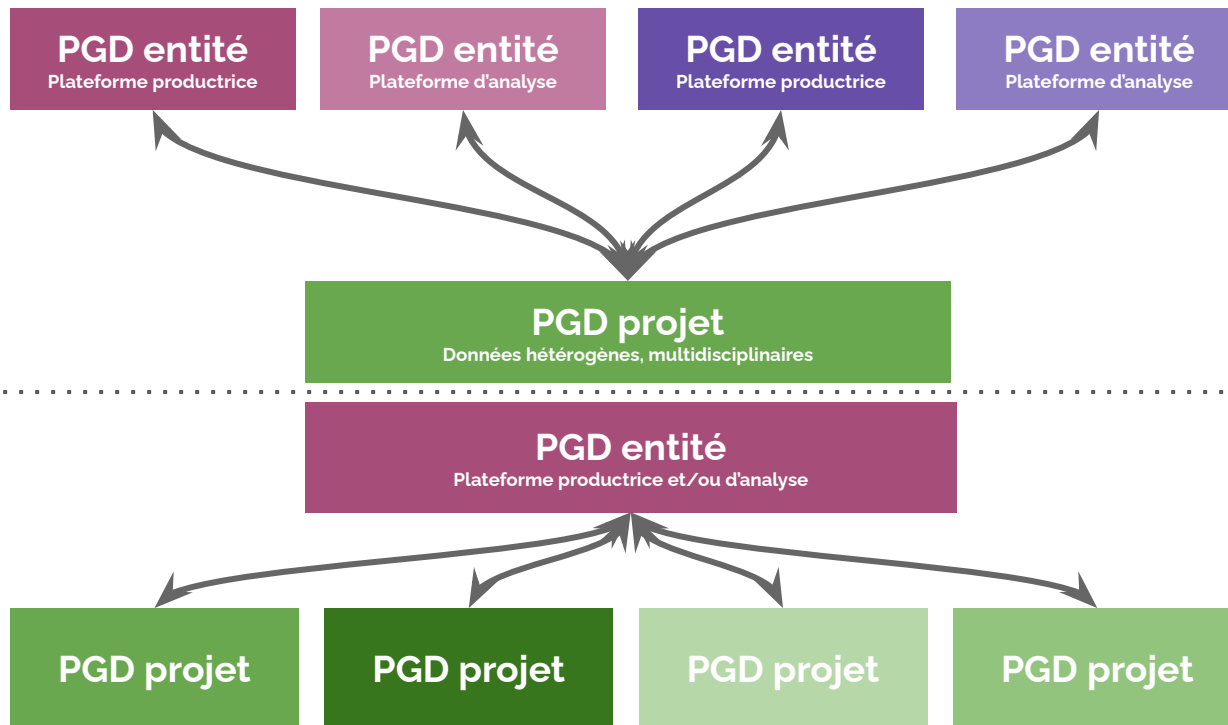
Mais dans la vraie vie... Rien n'est aussi simple





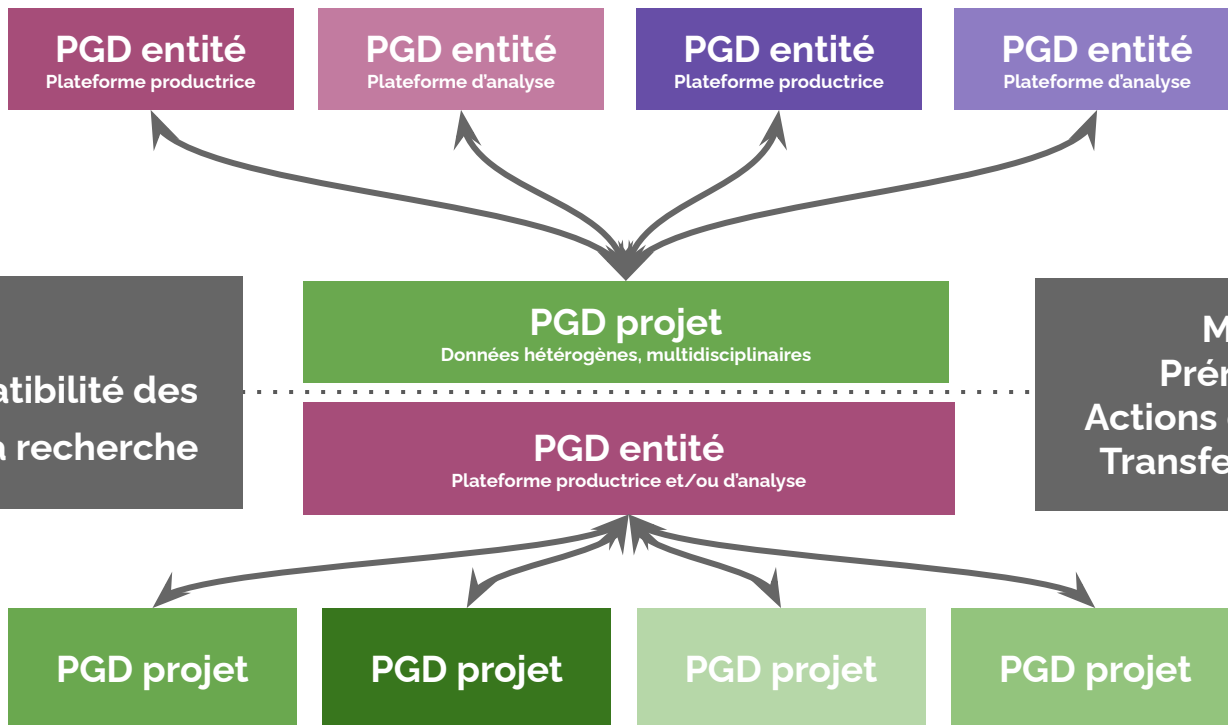
# Les PGD entités et les PGD projets en synergie

Mais dans la vraie vie... Rien n'est aussi simple



# Les PGD entités et les PGD projets en synergie

Mais dans la vraie vie... Rien n'est aussi simple



Qui réfléchit à tout ça ?



RESEARCH DATA ALLIANCE  
Structuration / réflexion  
internationale

  
**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE**

*Liberté  
Égalité  
Fraternité*

 **anr**<sup>®</sup>  
agence nationale  
de la recherche

CoSO

**Comité pour la science ouverte**



CoSIN

**Comité aux services et  
infrastructures numériques**

Réseau des ADAC

**Administrateur des données et  
codes sources de la recherche**

 **CENTRES DE RESSOURCES**  
Un accompagnement technique sur le dépôt de  
données sur Recherche Data Gov



 **RÉPUBLIQUE  
FRANÇAISE**  [recherche.data.gouv.fr](https://recherche.data.gouv.fr)

*Liberté  
Égalité  
Fraternité*

**Un écosystème au service du partage  
et de l'ouverture des données de recherche**



**CENTRES DE RÉFÉRENCES  
THÉMATIQUES**

Un spectre de compétences fondamentales relatives à  
la donnée et propre à chaque discipline



**ATELIERS DE LA DONNÉE**

Des experts de la donnée à disposition  
et à proximité des chercheurs



**MUSE ISDM**  
MONTPELLIER UNIVERSITÉ D'EXCELLENCE  
INSTITUT DE SCIENCE DES DONNÉES  
MONTPELLIER

**ECODOR**

 **Huma-Num**  
la TGR des humanités numériques

**Infrastructures  
nationales**

 **IFB**  
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

 **RARE**  
Ressources agronomiques  
pour la recherche

 **eosc**  **Huma-Num**  
Infrastructures  
européennes



 **EMPHASIS**  
EUROPEAN INFRASTRUCTURE FOR PLANT PHENOTYPING

*Niveau International*

*Niveau national*

*Niveau régional*

*Exemple à Montpellier*

**DipS INRAE**

 **DDOR** ...

Tutelles

Atelier à la CDGA, Grenoble, le 29 avril 2024



RESEARCH DATA ALLIANCE  
Structuration / réflexion internationale

Groupe de travail  
"PGD thématiques"



Infrastructures européennes



Réseau des  
"research data managers"  
<https://www.youtube.com/watch?v=S7HfUe1hWcg>

Niveau International  
Niveau national  
Niveau régional  
Exemple à Montpellier



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
Liberté  
Égalité  
Fraternité



CENTRES DE RESSOURCES  
Un accompagnement technique sur le dépôt de données sur Recherche Data Gov



Groupe de travail  
"PGD entités"

RÉPUBLIQUE FRANÇAISE  
recherche.data.gouv.fr  
Un écosystème au service du partage et de l'ouverture des données de recherche

ATELIERS DE LA DONNÉE  
Des experts de la donnée à disposition et à proximité des chercheurs



ECODOR

CENTRES DE RÉFÉRENCES THÉMATIQUES  
Un spectre de compétences fondamentales relatives à la donnée et propre à chaque discipline

- CoSO  
Comité pour la science ouverte
- CoSIN  
Comité aux services et infrastructures numériques
- Réseau des ADAC  
Administrateur des données, des algorithmes et des codes sources

Groupe de travail  
"Super-PGD de demain"



Infrastructures nationales



Groupe de travail  
"Modèles thématiques de PGD entités"



Tutelles

Et si on veut faire un PGD entité...  
Maintenant ?

# Quel modèle ?

**Modèle de PGD** : la série de questions à laquelle on répond lors de la rédaction.

Il existe de nombreux modèles de PGD :

- Spécifiques à des guichets de financement (ANR, H2020...)
- Spécifiques à des instituts (INRAE, CIRAD, CNRS...)
- Spécifiques à des centres (GENCI...)
  
- ❖ Certains modèles sont en partie structurés (Science Europe structuré)
- ❖ Certains modèles sont adaptés aux entités (Science Europe structuré, INRAE structure)
- ❖ Certains sont les deux
- ❖ Certains sont compatibles avec le *RDA DMP common standard*

**Le modèle entité n'est pas toujours très différent du modèle projet !**

# Quel modèle ?

Les modèle entité n'est pas toujours très différent du modèle projet !

Le projet / la structure	Obtention des données	Documentation des données	Stockage et sécurité des données
<ul style="list-style-type: none"> <li>Présentation du projet / de la structure</li> <li>Partenaires</li> <li>Politique en matière de gestion des données</li> </ul>	<ul style="list-style-type: none"> <li>Quelles données seront produites (type, format, volume...)</li> <li>Comment seront-elles produites ?</li> <li>Des données existantes seront-elles réutilisées ?</li> </ul>	<ul style="list-style-type: none"> <li>Quelle documentation et quelles métadonnées accompagneront les données ?</li> <li>Quelles procédures contrôlent la qualité des données ?</li> <li>Comment les données seront-elles organisées ?</li> </ul>	<ul style="list-style-type: none"> <li>Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?</li> </ul>
Exigences légales et éthiques	Accès et partage des données	Archivage des données	Responsabilités et ressources
<ul style="list-style-type: none"> <li>Droit applicable aux données produites</li> <li>Données à caractère personnel, respect du RGPD</li> <li>Données sensibles</li> <li>Prise en compte des questions éthiques</li> </ul>	<ul style="list-style-type: none"> <li>Quelles données seront partagées ou publiées ?</li> <li>Comment ? Avec qui ? Dans quel délai ?</li> <li>Quelle licence sera associée aux données ?</li> <li>Comment les données produites pourraient-elles être réutilisées ?</li> </ul>	<ul style="list-style-type: none"> <li>Comment les données seront-elles sélectionnées pour une préservation à long terme ? Y a t'il un plan d'archivage ?</li> </ul>	<ul style="list-style-type: none"> <li>Qui est responsable de la gestion des données ?</li> <li>Quelles sont les ressources dédiées à la gestion et à l'ouverture des données ?</li> </ul>



# Besoin de modèles thématiques

Préparer les PGD *machine actionable* avec des questions précises et adaptées

- ❖ Demande de nombreuses infrastructures nationales :

**France Génomique, France BioImaging, MetaboHUB, ProFi, ACF, EMBR-FR**

→ Groupe de travail

(Animation Paulette Lieby, Frédéric Brau)



- ❖ Groupe de travail, PGD thématiques (entité et projet)

(Animation Françoise Genova)



- ❖ DMP-OPIDoR, par le biais de formulaires spécialisés

# Quels outils ?

Nombreux outils sur la place internationale:



Deux focus stratégiques:

- ❖ DMP-OPIDoR:
  - outil favorisé au national
  - évolution constante (fonctionnalités *machine actionable*, ergonomie...)
  - nombreux modèles académiques, génériques
- ❖ DSW (Data Stewardship Wizard)
  - outil favorisé au niveau Européen (ELIXIR)
  - édition libre et collaborative des modèles

# Quels outils ?

Nombreux outils sur la place internationale:



Deux focus stratégiques:

- ❖ DMP-OPIDoR:
  - outil favorisé au national
  - nombreux modèles académiques, génériques



Besoin stratégique : interopérabilité



- ❖ DSW (Data Stewardship Wizard)
  - outil favorisé au niveau Européen (ELIXIR)
  - édition libre et collaborative des modèles



# Conclusion, sur les aspects théoriques

- ❖ Le PGD n'est pas un exercice à maturité
- ❖ Le PGD est un exercice clivant
- ❖ Le PGD incite à une étape de dialogue cruciale
- ❖ Le PGD recèle des informations “exclusives”
- ❖ Les différents PGD
  - Projets,
  - Plateformes,
  - Unités,
  - Établissements,

contiennent les mêmes questions, mais des réponses très différentes