Les Plans de Gestion des Données “structures”, ou “entités”

Paulette Lieby (IR) - cnrs

Jean-François Dufayard (chercheur) - cirad

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

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

QuaRES Association pour la Qualité en Recherche et Enseignement Supérieur

Mais qui êtes-vous ?

Mais qui êtes-vous ?

IFB: Centre de Référence Thématique en **biologie santé**

➔ proposer des référentiels pour la gestion des données

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

recherche.**data.gouv**.*fr*

Paulette Lieby: initialement issue des mathématiques fondamentales

- ➔ Animation des communautés imageries, multi-omiques
- ➔ Trames de PGD structures thématiques
- ➔ Research Data Alliance, ELIXIR

Jean-François Dufayard: initialement issu de l'algorithmique

- ➔ Co-coordination des relations entre infrastructures nationales
- ➔ Suivi du centre de référence thématique (Recherche Data Gouv)
- ➔ CoSO - collège des données

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

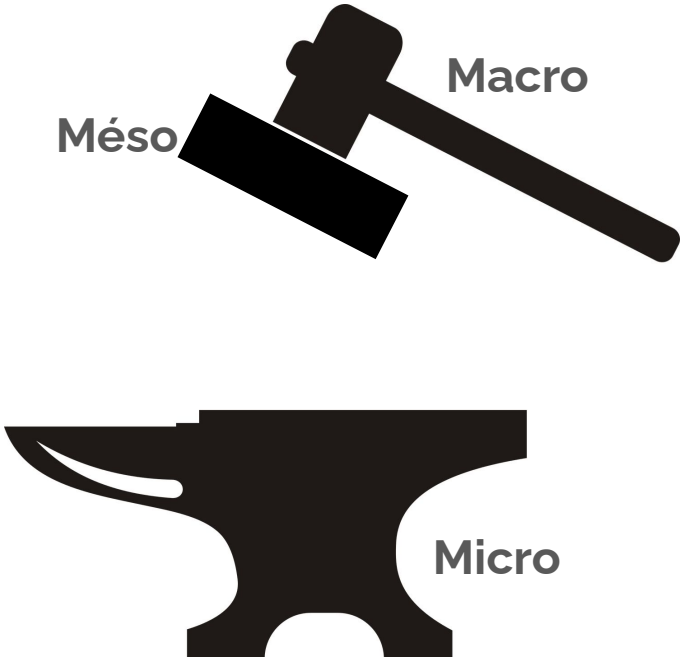
Macro

Méso gouvernance

➔ Le plan de gestion: notre degré de liberté

Micro

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… Le PGDDD le plan de gestion à **durée déterminée**

PGD structure ou entité : une stratégie de gestion, sur une structure pérenne (plateforme, laboratoire…) Le PGDDI le plan de gestion à **durée indéterminée**

PGD simple, PGD structuré

Attention au gag de vocabulaire

Structure ≠ Structuré

PGD simple/classique: ne contient que des réponses en texte libre

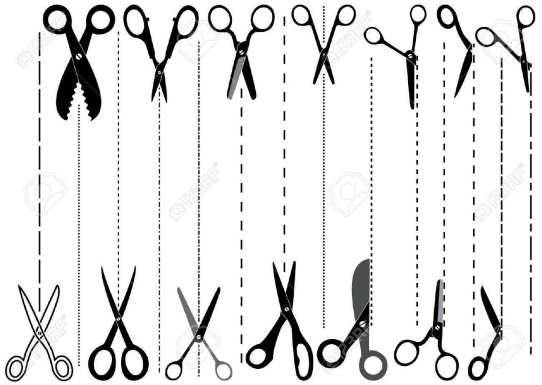
PGD structuré: une partie des réponses sont interprétables par une machine (chiffres, dates, choix multiples…)

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 produit de la recherche

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

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

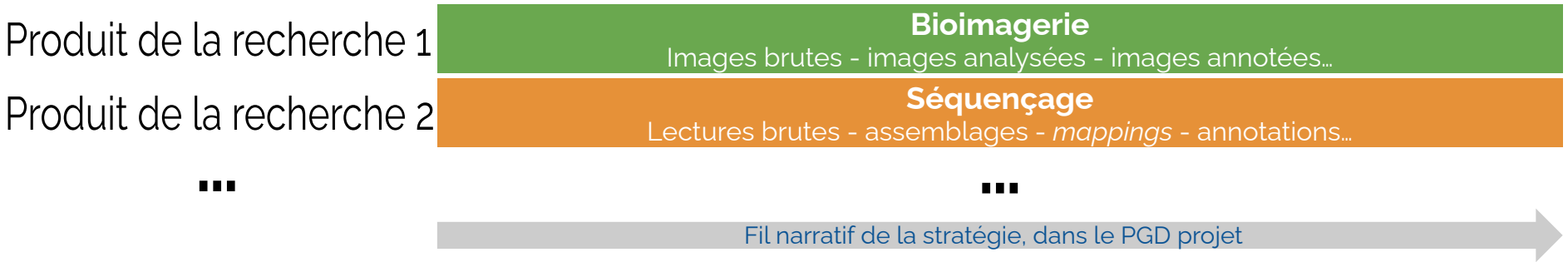
Le produit de la recherche - exemple projet

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

Par exemple, dans un même projet:

➔ Des données de bioimagerie

➔ Des données de séquençage

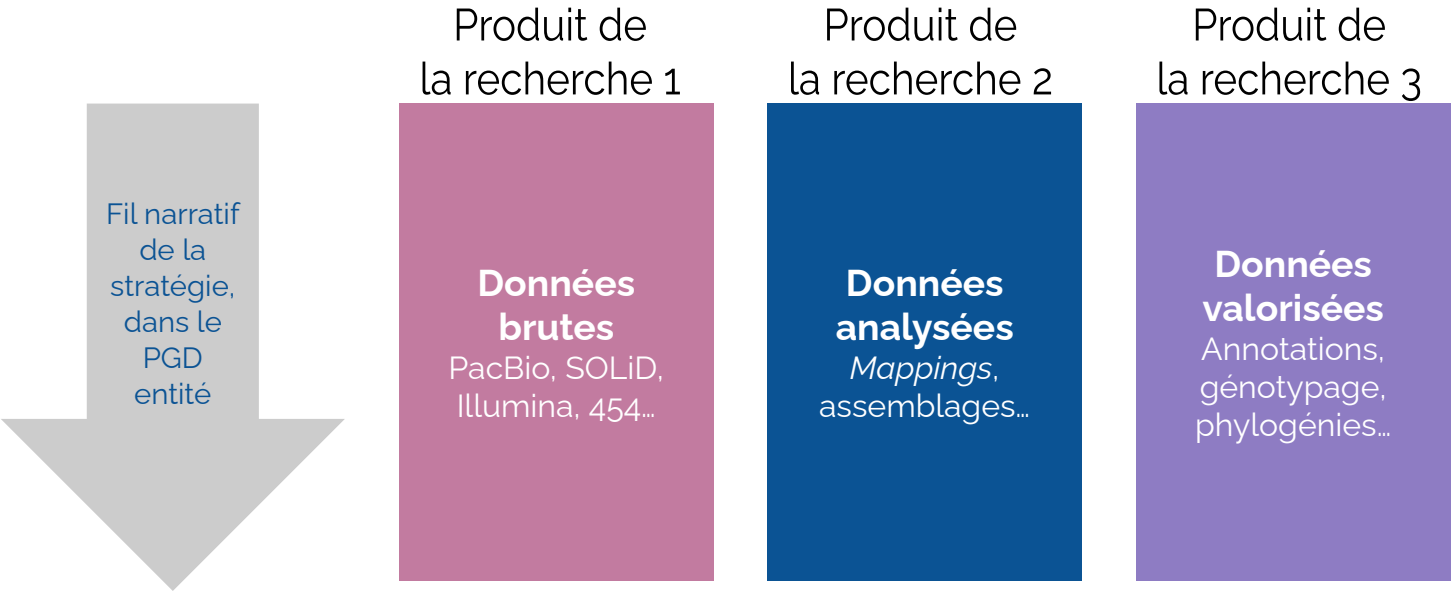


Le produit de la recherche - exemple entité / structure

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 (génomes annotés, génotypage, phylogénies…)



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

Les PGD entités et les PGD projets en synergie

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

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

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 ?

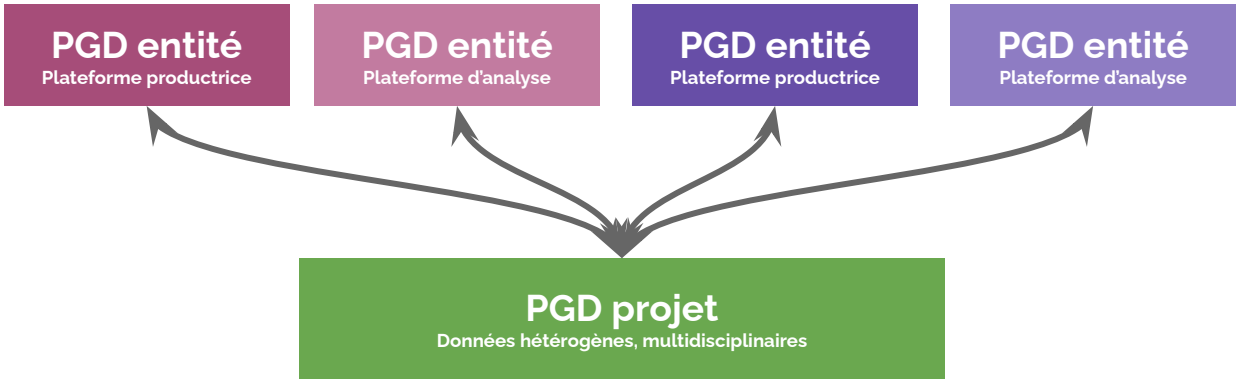
Quel matériel pour stocker et analyser les données ?

Quels tarifs ?

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

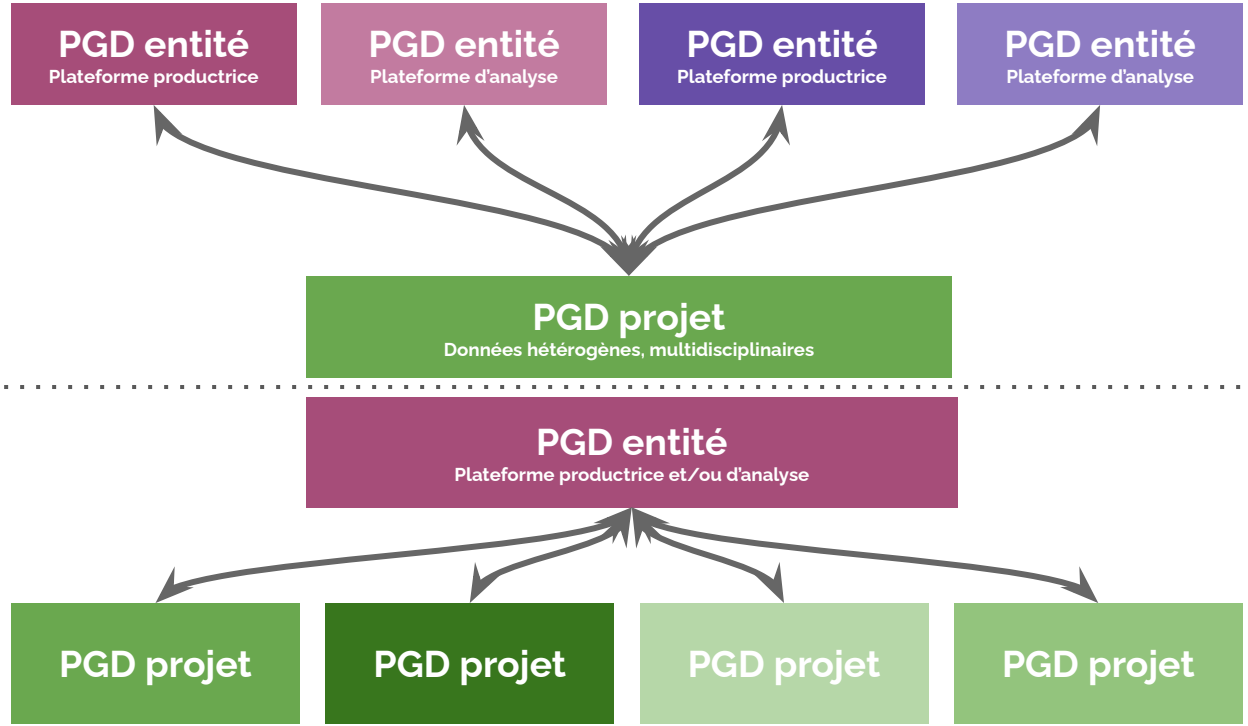
Les PGD entités et les PGD projets en synergie

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

PGD entité
Plateforme productrice

PGD entité
Plateforme d'analyse

PGD entité
Plateforme productrice

PGD entité
Plateforme d'analyse

PGD projet
Données hétérogènes, multidisciplinaires

PGD entité
Plateforme productrice et/ou d'analyse

PGD projet

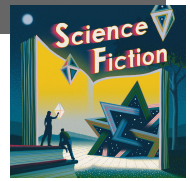
PGD projet

PGD projet

PGD projet

⚠ **Compatibilité des produits de la recherche**

Modularité
Préremplissage
Actions d'administration
Transferts de données



Qui réfléchit à tout ça ?

Structuration / réflexion internationale

eosc Huma-Num EMPHASIS ...

EUROPEAN INFRASTRUCTURE FOR PLANT PHENOTYPING

Infrastructures européennes

elixir

Niveau International

Niveau national

Niveau régional
Exemple à Montpellier

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

anr agence nationale de la recherche

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

OPIDoR

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

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

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

MUSE
MONTPELLIER UNIVERSITÉ D'EXCELLENCE

ISDM
INSTITUT DE SCIENCE DES DONNÉES MONTPELLIER

ECODOR

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

Huma-Num
la TGIR des humanités numériques

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE ...

Infrastructures nationales

RARe
Ressources agronomiques pour la recherche

Tutelles

DipSO INRAE

cnrs DDOR ...

Structuration / réflexion internationale

Groupe de travail “PGD thématiques”

eosc

Infrastructures européennes

Huma-Num

EMPHASIS

EUROPEAN INFRASTRUCTURE FOR PLANT PHENOTYPING

...

elixir

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

anr agence nationale de la recherche

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”

CENTRES DE RESSOURCES

Un accompagnement technique sur le dépôt de données sur Recherche Data Gouv

OPIDoR

Groupe de travail “PGD structures”

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

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

CENTRES DE RÉFÉRENCES THÉMATIQUES

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

MUSE

MONTPELLIER UNIVERSITÉ D’EXCELLENCE

ISDM

INSTITUT DE SCIENCE DES DONNÉES MONTPELLIER

ECODOR

Huma-Num

la TGIR des humanités numériques

IFB

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

...

Infrastructures nationales

RARe

Ressources agronomiques pour la recherche

Groupe de travail “Modèles thématiques de PGD entités”

Tutelles

DipSO INRAE

cnrs DDOR

...

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…)

Tous compatibles avec le *RDA DMP common standard*

- ❖ Certains modèles sont en partie structurés (Science Europe structuré)
- ❖ Certains modèles sont adaptés aux entités (INRAE structure)
- ❖ Certains sont les deux

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

Quel modèle ?

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

Le projet / la structure	Obtention des données	Documentation des données	Stockage et sécurité des données
• Présentation du projet / de la structure • Partenaires • Politique en matière de gestion des données	• Quelles données seront produites (type, format, volume…) • Comment seront-elles produites ? • Des données existantes seront-elles réutilisées ?	• Quelle documentation et quelles métadonnées accompagneront les données ? • Quelles procédures contrôlent la qualité des données ? • Comment les données seront-elles organisées ?	• Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?
Exigences légales et éthiques	**Accès et partage des données**	**Archivage des données**	**Responsabilités et ressources**
• Droit applicable aux données produites • Données à caractère personnel, respect du RGPD • Données sensibles • Prise en compte des questions éthiques	• Quelles données seront partagées ou publiées ? • Comment ? Avec qui ? Dans quel délai ? • Quelle licence sera associée aux données ? • Comment les données produites pourraient-elles être réutilisées ?	• Comment les données seront-elles sélectionnées pour une préservation à long terme ? Y a t'il un plan d'archivage ?	• Qui est responsable de la gestion des données ? • Quelles sont les ressources dédiées à la gestion et à l'ouverture des données ?

Besoin de modèles thématiques

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

- Demande de nombreuses infrastructures nationales:

  France Génomique, France BioImaging, MetaboHUB, ProFi…

  Besoin de trames thématiques pour leurs PGD de plateformes

  ➔ Groupe de travail

  (Animation Paulette Lieby, Frédéric Brau)

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

  (Animation Françoise Genova)

Quels outils ?

Nombreux outils sur la place internationale:

argos

DMPTool + DMP ONLINE = roadmap

DSW DATA STEWARDSHIP WIZARD

DMP OPIDoR

Deux focus stratégiques:

- DMP-OPIDOR:
  - outil favorisé au national
  - 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

Interopérabilité - besoin stratégique

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

Démonstration en mode atelier