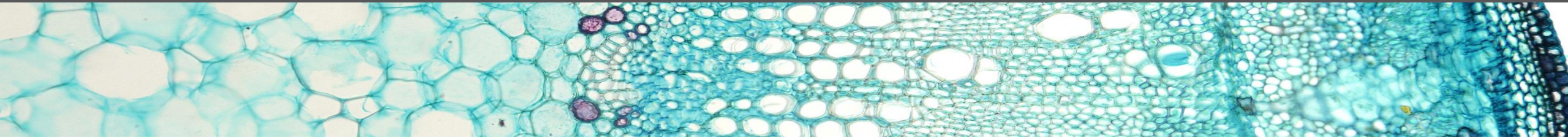


Les plateformes BioCampus : leur gestion des données et leur PGD



Oana Vigy

oana.vigy@fpp.cnrs.fr

Bi  Campus
Montpellier



Inserm





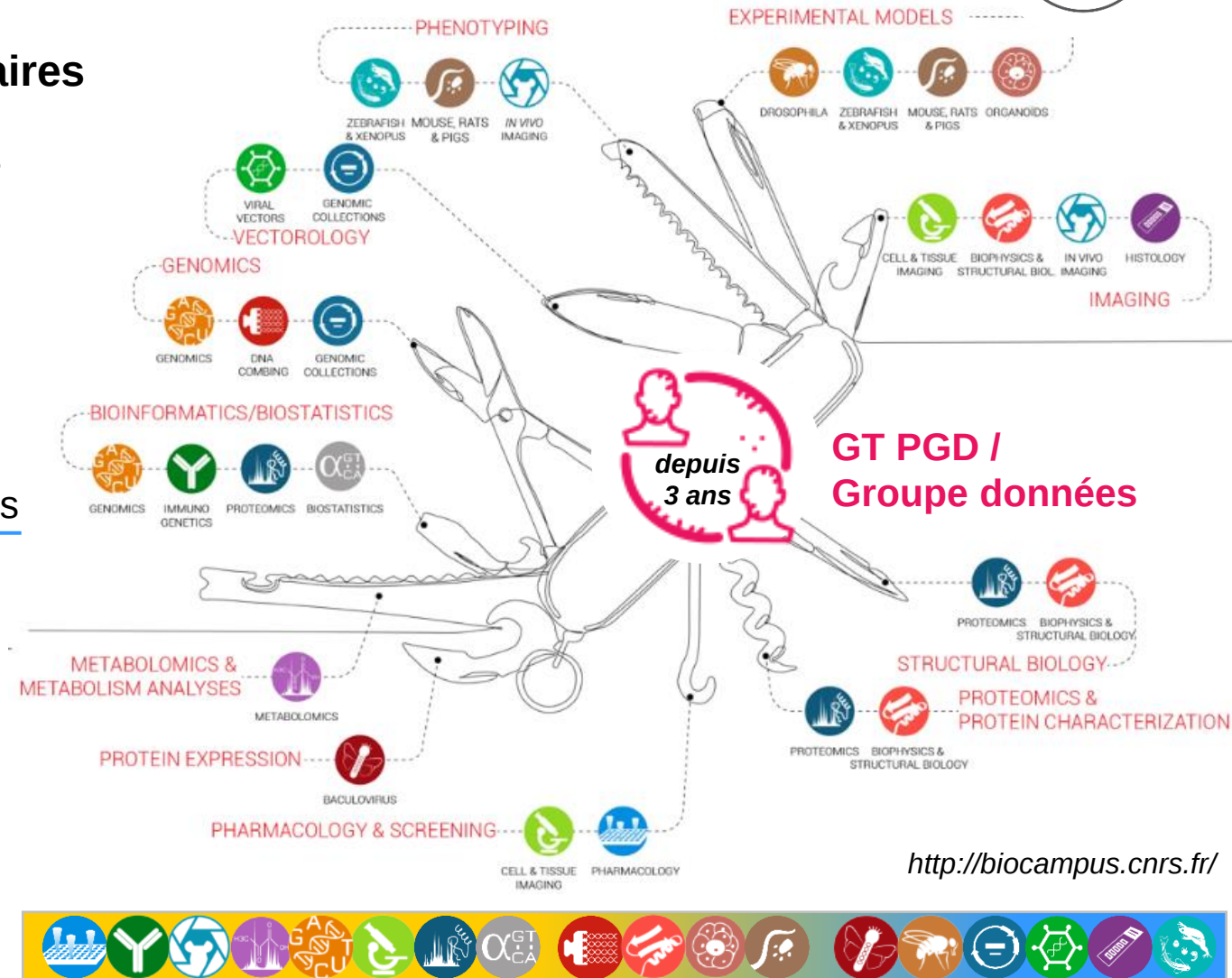
19 plateformes technologiques multi-disciplinaires

- Mono et multi-plateaux répartis sur différents campus
- Multi-tutelles
- Labellisation et certification
- PF acquisition / production de données
(avec/sans mise à disposition d'équipement)
- PF production / transformation de produits biologiques
(gestion de métadonnées)

BiCampus
Montpellier

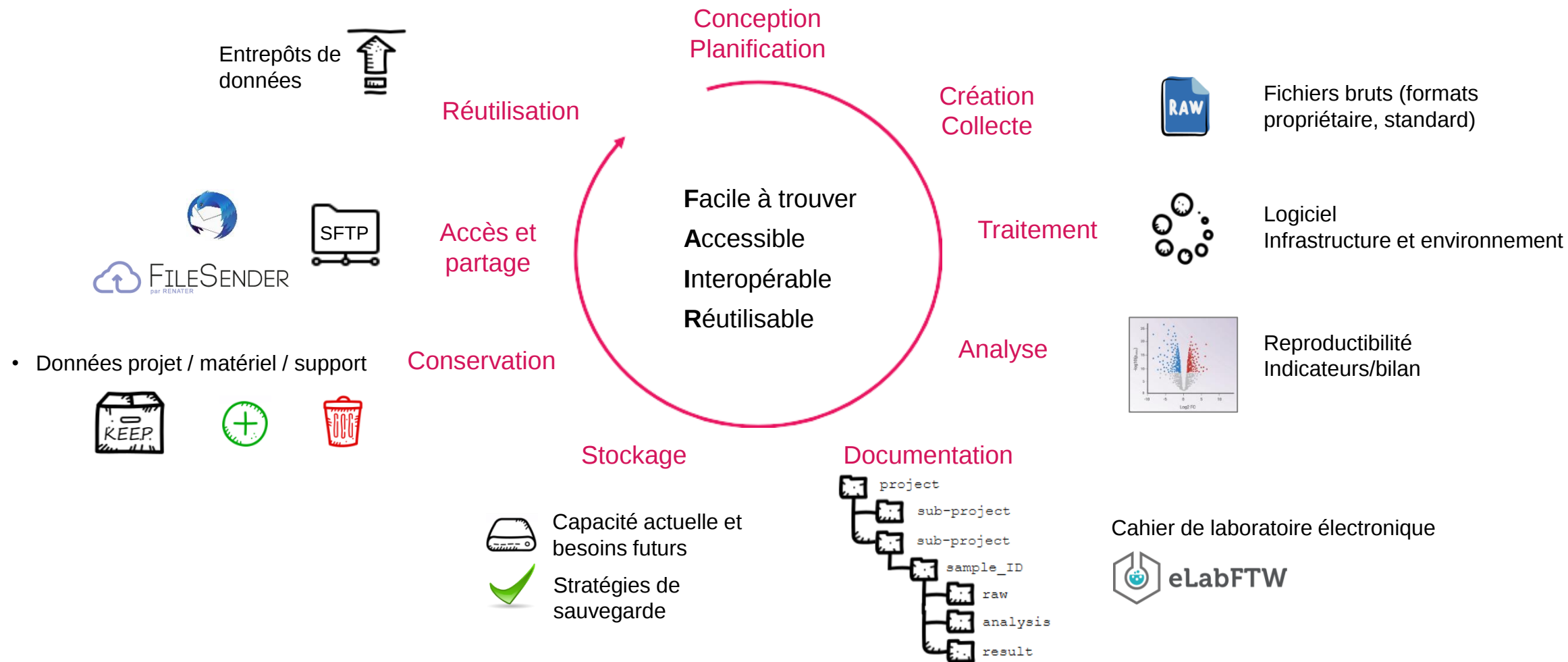


Inserm





Exemple d'une plateforme d'acquisition et d'analyse de données





Les objectifs du GT

- Formaliser dans 1 document la gestion des données réalisée par chacune des plateformes
- Capitaliser les apports des plateformes qui sont engagées dans la rédaction d'un PGD

2022

Participation active de $\pm$ les 19 plateformes, réunion ~ tous les trimestres

2025

Un **PGD** ~~de structure~~ d'entité, c'est à dire

- Décrire comment les données sont obtenues
- Garantir la compréhension des données
- Clarifier le cadre juridique et éthique
- Assurer un stockage adapté
- Spécifier les modalités de partage
- Définir le rôle de chacun



L'organisation du GT

- Animer une démarche incitative avec des volontaires ☺
- Se former pour **démystifier la thématique**

2022

2023

2024

2025

Formation sur les principes FAIR pour la gestion des données d'une plateforme IBiSA



- Sessions théorique (1 j) + pratique (2 j)
- 3 plateformes représentées



<https://moodle.france-bioinformatique.fr/course/view.php?id=16>

Quelques conseils pour bien commencer

- Rédaction du PGD
 - Identifier un « référent PGD », la personne qui va fédérer/animer le groupe qui remplit le PGD
 - Bien identifier les « produits de recherche » => la base
 - Commencer par un état des lieux, les améliorations arriveront dans un second temps
 - Trouver la bonne granularité, doit être utile et maintenable (équilibre exhaustivité/généralité)
 - Se fixer des deadlines avec des objectifs raisonnables, travailler étape par étape
 - Travail d'équipe, mais proposer une ébauche et non une page blanche aux collègues
- OPIDoR (<https://dmp.opidor.fr/>)
 - Trame = aide à la construction des PGD, aide à se poser les bonnes questions
 - Attention aux questions qui semblent redondantes
 - Réflexion sur les questions qui ne semblent pas claires
 - Trame héritée des PGD de Projets, donc très orientée projet => remplacer « projet » par « structure » pour certains champs, ignorer d'autres champs
 - **A la fin**, ne pas hésiter à exporter le document pour l'enrichir (ajout d'items, schémas, tableaux ...)

=> Le plus dur c'est de se lancer !

Problématique : quelle **structuration**, quelle articulation ?

PGD de plateforme



Point commun :
stratégie de gestion
des données

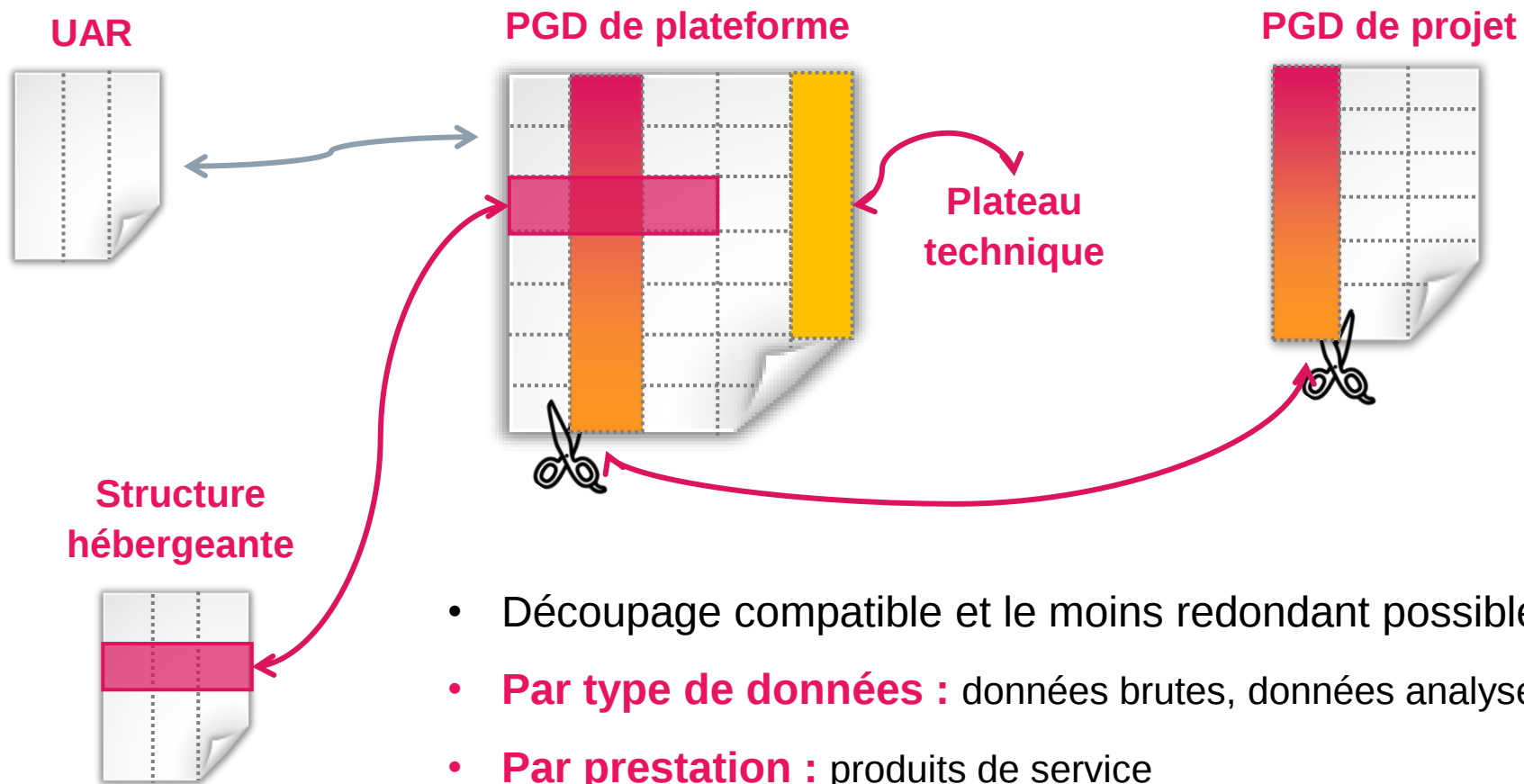


PGD de projet



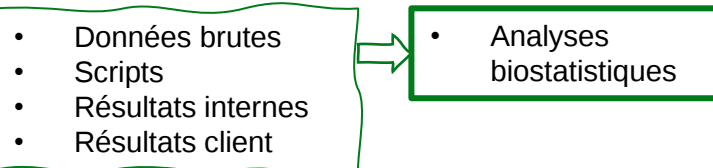
- Réfléchir sur les **produits de service / recherche** qui structurent le PGD
(*ensemble de données gérées de la même façon*)
- Document lisible et adapté aux mises à jour

Problématique : quelle structuration, quelle **articulation** ?

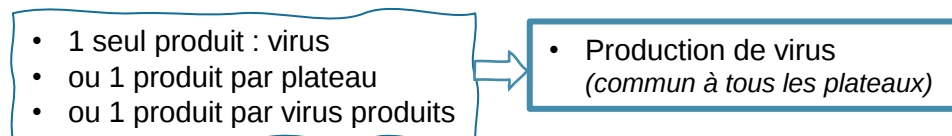
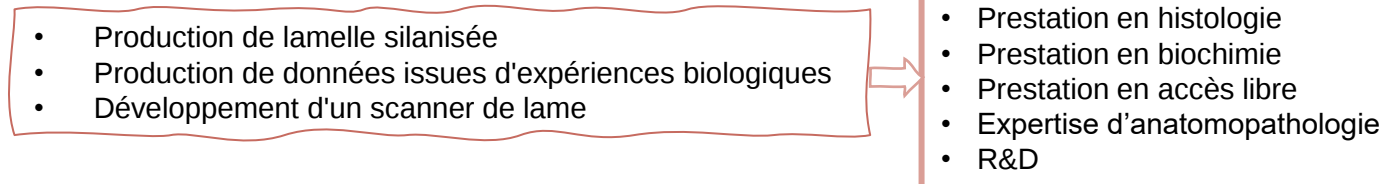
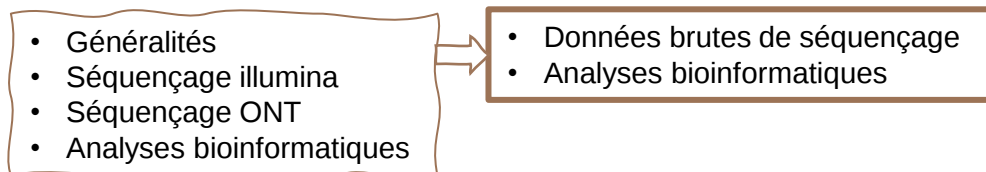


- Découpage compatible et le moins redondant possible
- **Par type de données** : données brutes, données analysées, données de résultats
- **Par prestation** : produits de service

Produits de recherche ... Produits de service



- Données d'annotations, captation des données, production d'ontologies
- Mise à disposition d'outils d'analyse
- Élaboration et maintenance d'outils



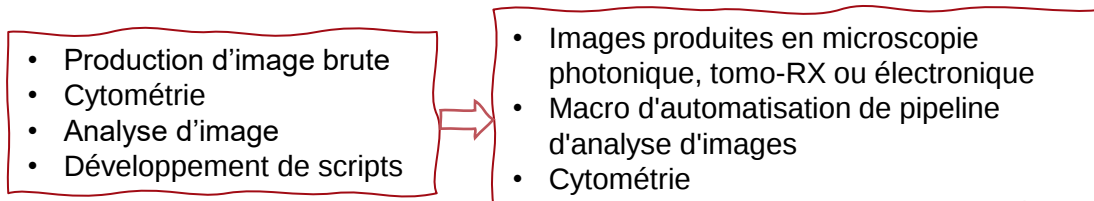
- Consommation d'oxygène
 - Potentiel de membrane mitochondrial
 - Production d'H₂O₂
- 1 site



- Identification de protéines en HRMS
 - Dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS
 - Dosage de protéines par MRM
 - Dosage ELISA (MSD, Simoa, Ella) de molécules
- 1 site

- ↓
- Analyse bottom-up
 - Analyse ciblée
 - Analyse protéomique structurale
 - Dosage (Elisa)
 - Interaction (SPR)
 - Nano-DSF
- n sites

- Données analytiques brutes générées par les logiciels d'acquisitions
- Données retraitées



- Prestation accès-libre
- Prestation de service
- Recherche & Développement

- Production d'images
- Analyse

S'appuyer sur les outils ... et les adapter



Problématique (*bis*) : tenir compte de l'existant, des évolutions et des échéances

On ne dispose pas aujourd'hui de consensus clair sur le **template du PGD à employer**

Outils d'aide à la rédaction



- Communauté d'utilisateurs
- *Machine actionnable* et interopérabilité
- Evolution des modèles et conséquences sur son PGD
- Versionnage

De multiples modèles



- Modèle **entité**
- Modèle projet



- Modèle **entité**



- Modèle **entité**

Recommandé !



Modèle « structuré »
(*machine actionnable*)



- Modèle projet

S'appuyer sur les outils ... **et les adapter**



Problématique (*bis*) : tenir compte de l'existant, des évolutions et des échéances

Outils d'aide à la rédaction



De multiples modèles



- Modèle **entité**
- Modèle projet



- Modèle **entité**



- Modèle **entité**



- Modèle projet



- Elaboration d'un modèle de PGD entité multi-omiques

S'appuyer sur les outils ... et les adapter



Travail collaboratif sur un modèle de PGD de structure multi-omique

5.1 Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

• Décrire les besoins de stockage

B I U A

DMP OPIDoR

• Volume estimé des données

• Unité

Sélectionnez une valeur de la liste ou saisissez

Basé sur le modèle ScienceEurope...

... adapté pour et avec l'aide des communautés multi-omiques

- ▼ Si connue, estimation de la taille des données
- ▼ Éléments
- ▼ Comment décrivez-vous la volumétrie ?
- ▼ Description de la volumétrie par élément
- ▼ Éléments
- S'agit-il de jeux d'images ?
 - S'agit-il de jeux de données cytométrie ?
 - S'agit-il de jeux de données génomiques ?
 - S'agit-il de jeux de données métabolomiques ?
 - ▼ S'agit-il de jeux de données protéomiques ?
 - De quel élément parle-t-on ?
 - Quelle est la volumétrie de l'élément ?
 - Dans quelle unité ?
 - Pouvez-vous estimer le nombre d'éléments ?
 - S'agit-il de données autres ?



✓ III.2.a.4.a.2.a.2.a.1.b.1.a.5.b.1

De quel élément parle-t-on ?

Proteomique

Quelques exemples d'élément :

- un run d'acquisition par temps d'acquisition (à spécifier)
- une analyse d'un échantillon global : avec/sans fractionnement, nombre de réplicats, logiciel d'analyse

☒ Désirable: Avant soumission du PGD

1 | run d'acquisition par temps d'acquisition : 1h

Supprimer la réponse

Réponse à il y a moins d'une minute par Oana Vigy.

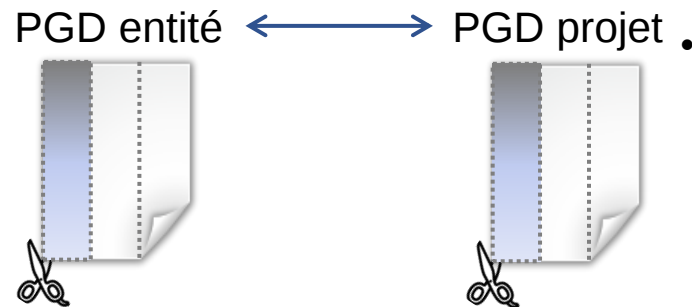
<https://dsw.france-bioinformatique.fr/wizard/>

Présenter la **politique de gestion des données**

- Mise en avant (et à plat) de la gestion des données
- Amélioration des pratiques (*démarche qualité, point d'attention lors des audits*)
- Recommandations sur les bonnes pratiques
- Travail collaboratif et structurant entre les différents acteurs identifiés
- Identification de la bonne donnée et du bon temps de conservation (*gain financier et écologique*)
- ...

Décrire les données produites dans le cadre d'un **service à la recherche**

- Document de référence auquel peuvent se rapporter les utilisateurs (*gain de temps*)
- Rend visible l'expertise et l'implication de la plateforme dans la gestion des données
- Dépôt de données obligatoire dans la plupart des publications (*gain de temps*)
- Gain de temps (à venir) avec l'interopérabilité entre PGD projet et PGD entité
- ...





Les avancées ...

- Mise en avant (et à plat) de la gestion des données
- Travail collaboratif et structurant
- Valorisation de notre expertise

Avoir une **réflexion** sur le cycle de vie des données qu'on génère et de réfléchir à comment **améliorer** nos procédures.

C'est clairement pas une priorité surtout qu'**aucune équipe de recherche nous le demande**

2022

Progression entre découverte, réflexion jusqu'à quelques PGD finalisés

2025



Les avancées ...

- Mise en avant (et à plat) de la gestion des données
- Travail collaboratif et structurant
- Valorisation de notre expertise

Avoir une **réflexion** sur le cycle de vie des données qu'on génère et de réfléchir à comment **améliorer** nos procédures.

C'est clairement pas une priorité surtout qu'**aucune équipe de recherche nous le demande**

2022

Progression entre découverte, réflexion jusqu'à quelques PGD finalisés

2025

... malgré les difficultés rencontrées

- Prioriser ce travail parmi tant d'autres
- Définir les « **produits de recherche** » pour organiser au mieux le PGD (approche top-down)
- Utiliser les outils à notre disposition en toute connaissance de cause
- Trouver le bon niveau de détail tout en évitant les redondances (documentation qualité, etc.)
- Passer du plateau à la plateforme (multi-sites)

Trouver le **temps** pour la rédaction et la mise en place des actions décrites dans le PGD

Le principal problème rencontré est l'**hétérogénéité** des données correspondant à 8 plateaux très différents

Remerciements – On continue !



**Equipe du service Science Ouverte
et accompagnement à la recherche**



**Paulette Lieby et tout le
GT2 IS1 du projet Mudis4LS**