

Bienvenue



FAIR

by IFB



Fred de Lamotte - Montpellier
<https://orcid.org/0000-0003-4234-1172>



INRAE

**IFB**
INSTITUT FRANÇAIS DE BIOINFORMATIQUE



Session 1 Séquence 5

Fred de Lamotte - Montpellier

<https://orcid.org/0000-0003-4234-1172>

Jean-François Dufayard

<https://orcid.org/0000-0002-7427-6822>

Paulette Lieby

<https://orcid.org/0000-0002-9289-9652>

Donc, un PGD

Un plan de gestion de données

Ou un DMP : data management plan



Le plan de gestion des données

25 minutes

POUR QUOI FAIRE ?

- **Plan** : on planifie (donc on anticipe)
- **Gestion** : on gère, on fait fructifier (on commence déjà par ne plus perdre)
- **Données** : Data is the new oil, the new soil

GÉRER LES DONNÉES : COMMENT ?

- Une approche pragmatique
 - Simple à comprendre
 - Simple à mettre en place
 - Simple à évaluer
 - Simple à faire évoluer

- Donc le Plan de Gestion des Données
 - Actuellement... un document texte (en attendant mieux)

LES OBJECTIFS DU PGD

- Assurer la reproductibilité des expériences
 - Décrire comment les données sont obtenues
- Respecter le droit et les personnes
 - Clarifier le cadre juridique et éthique
- Permettre la réutilisation des données
 - Garantir la compréhension des données
- Éviter les pertes de données
 - Assurer un stockage adapté
- Établir le rôle de chacun
 - Définir les responsabilités
- Clarifier les droits de réutilisation
 - Spécifier les modalités de partage

LES OBJECTIFS DU PGD

- Assurer la reproductibilité des expériences
 - Décrire les données

<https://dmp.opidor.fr/>

IFB_training WGS RNAseq Variant_calling

Quels méthodes et outils sont utilisés pour acquérir et traiter les données ? Précisez les différents formats dans lesquels les données seront disponibles aux différentes phases de la recherche

B I    

- Format : fastq, outil : séquenceur Illumina HiSeq 3500

- Format : fasta, outil : assembleur SOAP de novo

- Format : gff3, outil : Eugene

Enregistrer

Répondu 1 month ago par helene.chiapello@inrae.fr

Commentaires (1)

Vérifier avec Gautier si le mode opératoire de l'Illumina est stabilisé

frederic.de-lamotte@inrae.fr 1 hour ago

Modifier Supprimer

Ajouter un commentaire à partager avec les collaborateurs

B I    

LES OBJECTIFS DU PGD

- Respecter le droit et les personnes
 - Clarifier le cadre juridique et éthique

IFB_training WGS RNAseq Variant_calling

Les réponses de cette section sont communes à tous les produits de recherche

Qui détiendra les droits sur les données et les autres informations créées lors du projet ?

Faire attention quand un partenaire privé amène des données dans le projet.

B *I*    

par défaut INRAE sinon voir convention

Enregistrer

Répondu 1 month ago par helene.chiapello@inrae.fr

Commentaires (1)

- j'ai un souci : depuis la loi LRN les données sont ouvertes par défaut, sauf exceptions (données personnelles, sensibles, code, autres...). Est-il encore opportun de parler de "propriété intellectuelle" dans ce cas
- je me réfère là à un commentaire de Lionel Maurel "La question de la propriété des données n'est pas forcément le bon angle d'attaque depuis LPRN (et principe d'ouverture par défaut). Aujourd'hui savoir qui est le propriétaire des données n'est plus souvent. Il s'agit de savoir si on est dans le principe de l'ouverture par défaut ou bien dans un cas d'exceptions." -- compte rendu atelier <https://mate-shs.cnrs.fr/actions/tutomate/tuto25-propriete-donnees-lionel-maurel/>
- en fait, je crois qu'ici cela demandera à terme une re-

LES OBJECTIFS DU PGD

- Permettre la réutilisation des données
 - Garantir la compréhension des données

Brève présentation des données générées, collectées ou réutilisées :

- Mode d'obtention, origine, type, nature et périmètre thématique des données
- Publications associées

B *I*    

Mode d'obtention : sortie d'un workflow de variant calling

Origine : vitis vinifera (la vigne)

Type : variants au format VCF

Enregistrer

INRAE Exemple de réponse +

Répondu 1 month ago par helene.chiapello@inrae.fr

Commentaires

Ajouter un commentaire à partager avec les collaborateurs

B *I*    

Merci de rajouter un lien vers le workflow

Enregistrer

LES OBJECTIFS DU PGD

- Éviter les pertes de données
 - Assurer un stockage adapté

IFB_training WGS RNAseq Variant_calling

Les réponses de cette section sont communes à tous les produits de recherche

Stockage : Quels seront les supports utilisés pour les données au cours du projet ?

B *I*    

Préciser l'infrastructure et la plateforme bioinformatique utilisée (IFB, Southgreen, migale,...)

Enregistrer

INRAE Exemple de réponse +

Répondu 1 month ago par helene.chiapello@inrae.fr

Commentaires

Ajouter un commentaire à partager avec les collaborateurs

B *I*    

Jeff doit réserver 2 To avant le début du projet !

Enregistrer

LES OBJECTIFS DU PGD

- Établir le rôle de chacun
 - Définir les responsabilités

Gérer les collaborateurs

Inviter des personnes à lire, modifier ou administrer votre plan. Les invités recevront une notification par courriel indiquant qu'ils ont accès à ce plan.

Adresse courriel	Permissions
thomas.denecker@france-bioinformatique.fr	Éditeur  Supprimer
seilerj@igbmc.fr	Éditeur  Supprimer
paulette.lieby@france-bioinformatique.fr	Éditeur  Supprimer
gautier.sarah@inrae.fr	Copropriétaire  Supprimer
frederic.de-lamotte@inrae.fr	Copropriétaire
helene.chiapello@inrae.fr	Propriétaire

Inviter des collaborateurs

* Courriel

* Permissions

- Co-propriétaire: peut modifier les détails du projet, changer la visibilité et ajouter des collaborateurs.
- Éditeur: peut commenter et effectuer des changements
- Lecture seule: peut voir et commenter, mais ne peut pas faire de modifications

UN PGD – PLUSIEURS VERSIONS

- C'est un document évolutif
 - au moins 3 versions :
- V1 à 6 mois
 - C'est un livrable pour H2020 et ANR
- V2 à mi-parcours
 - Quand la connaissance des données progresse
- V3 en fin de projet
 - Quand le flux de données se calme et qu'on a une bonne vision globale
 - Le confinement est un moment idéal pour ce genre de ménage !

MODÈLES POUR LES PGD

Modèle de PGD = une liste de sections, questions, et informations à remplir pour rédiger un PGD

Modèles existants:

- RDA - Research Data Alliance - un “modèle originel”
- Des modèles spécifiques par tutelles (INRAe, CEA, CIRAD, universités, grandes écoles...)
- Des modèles par grands guichets (HORIZON 2020 UE, ANR...)
- Des modèles pour certains centres de calcul/stockage (IN2P3)

Peu / pas de modèles spécifiques à des types de données

OUTILS POUR LES PGD

Systemes pour rédiger des DMPs

- DMP OPIDOR - solution nationale



DS Wizard

20210224_exemple_canevas_IFB_bioimage

Questionnaire | TODOs | Metrics | Preview | Documents | Settings

Chapters

- I. Préface ✓
- II. Introduction 1
- III. Informations générales 24
- IV. Données de la recherche 93**

1.a.2.a.2.a.3 Quelles mesures de contrôle de la qualité sont prises pour ce jeu de données ?

Ceci concerne l'acquisition des images. Voir le premier white paper de <https://quarep.org/>, <https://arxiv.org/pdf/2101.09153.pdf>. La liste d'options présentée ici est en anglais : ne devrait-on pas traduire cela pour être consistant/cohérent avec le reste de ce document ? (quoique les modes d'acquisition sont bien donnés en anglais...)

- a. Illumination power
- b. Detection system performance
- c. Field of view uniformity.flatness
- d. Chromatic aberrations
- e. Lateral and axial resolution
- f. Image quality
- g. Commentaires ☰

1.a.2.a.2.a.4 Des versions différentes du jeu de données sont-elles créées ?



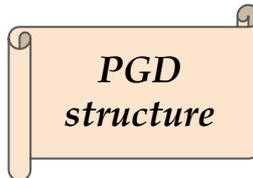
ET SI NOUS SORTIONS DE LA PRÉHISTOIRE ?

Dichotomie

Structure

Plateforme
Infrastructure
Service...

PGD à durée indéterminé
PGD plus générique



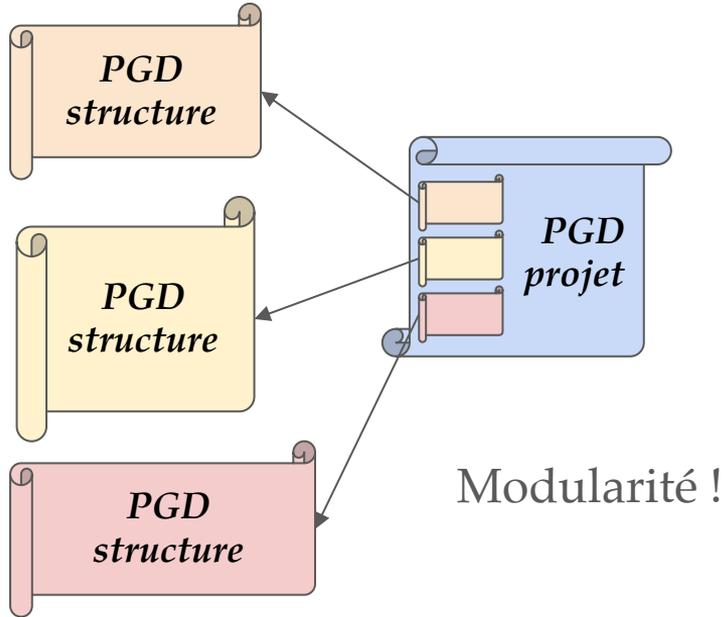
Projet

Projet d'imagerie
Projet multi-omique
Projet plus complexe ...

PGD à durée déterminé
PGD plus spécifique

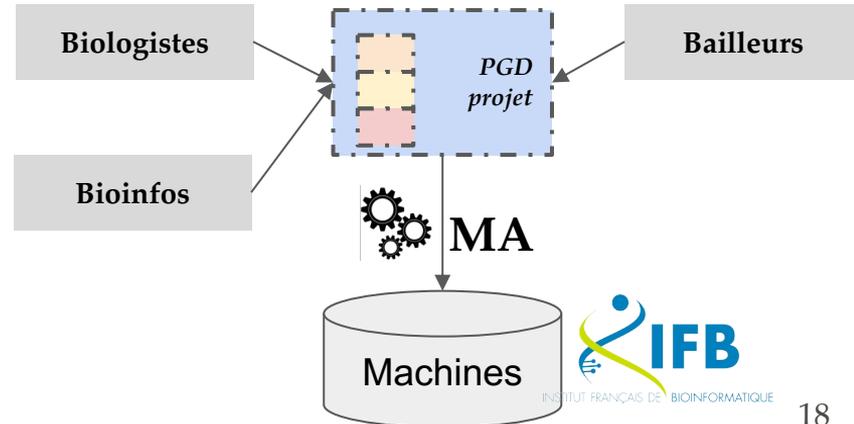


ET SI NOUS SORTIONS DE LA PRÉHISTOIRE ?



Machine actionable DMP: un plan de gestion lisible par les machines

Objectif : faire du Plan de gestion des Données un outil de configuration des environnements des infrastructures



POUR LA SUITE

Aller sur <https://dmp.opidor.fr>

Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.



Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers



Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données



Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche



Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

Rejoignez la communauté des utilisateurs de DMP OPIDoR

Créez un compte, connectez-vous et laissez-vous guider !

[Découvrez DMP OPIDoR](#)

The screenshot shows the login interface of the DMP OPIDoR website. At the top, there are two tabs: 'Connexion' (selected) and 'S'inscrire'. Below the tabs, there are two main sections: 'Accès institutionnel' and 'Accès individuel'. Under 'Accès institutionnel', there is a prominent red button labeled 'Connexion'. Under 'Accès individuel', there are two required fields: '* Courriel' and '* Mot de passe', each with a text input box. Below these fields, there is a link for 'Mot de passe oublié ?' and a checkbox for 'Se souvenir de moi'. At the bottom of the 'Accès individuel' section, there is another red button labeled 'Connexion'.

Titre du projet



projet de test, d'entraînement ou à des fins de formation

Choisissez un modèle

Vous pouvez choisir soit un modèle fourni par votre organisme soit par un autre organisme, ou un modèle financeur. Le modèle par défaut est **Science Europe - DMP template (english)**.

[Retrouvez la liste des modèles disponibles](#)

INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement (Votre organisme)	Autre organisme	Financier
<p data-bbox="195 794 1132 825">Plusieurs modèles sont disponibles, lequel souhaitez-vous utiliser ?</p> <div data-bbox="195 838 929 882"><input type="text" value="INRAE - General project template"/></div> <p data-bbox="195 887 649 909">Veuillez sélectionner un modèle dans la liste.</p> <div data-bbox="1089 942 1709 986">Créer un plan Suivant Utiliser le modèle par défaut</div>		



 Gestion et partage des données scientifiques Contact (Pôle Données de la recherche de l'IST) Portail Data INRAE

IFB_Training_plant

[Renseignements sur le projet](#)[Produits de recherche](#)[Modèle choisi](#)[Rédiger](#)[Partager](#)[Demande d'assistance conseil](#)[Télécharger](#)

Gérer les collaborateurs

Inviter des personnes à lire, modifier ou administrer votre plan. Les invités recevront une notification par courriel indiquant qu'ils ont accès à ce plan.

Adresse courriel	Permissions
lionel.maurel@cnr.fr	<input type="text" value="Éditeur"/> Supprimer

Inviter des collaborateurs

*** Courriel**

*** Permissions**

- Co-propriétaire: peut modifier les détails du projet, changer la visibilité et ajouter des collaborateurs.
- Editeur: peut commenter et effectuer des changements
- Lecture seule: peut voir et commenter, mais ne peut pas faire de modifications

[Sauvegarder](#)