



Session 1 Séquence 5

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



Donc, un PGD

Un plan de gestion de données

Ou un DMP : data management plan



Le plan de gestion des données

25 minutes

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



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

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 [List] [List] [Link] [Table]

- 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 [List] [List] [Link] [Table]

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>I</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>I</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

Les objectifs du PGD

- Clarifier les droits de réutilisation
 - Spécifier les modalités de partage

Sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ? Sinon, quelles procédures seront mises en place pour la conservation à long terme ?

Il s'agit ici de plateformes d'archivage pérennes destinées à pérenniser les données, comme le C.I.N.E.S. Les entrepôts de données ne possèdent, à quelques exceptions près, pas cette possibilité.

B	<i>I</i>	☰	☰	🔗	📄
<p>En biologie ce sont jusqu'à maintenant les banques internationales (ENA, ArrayExpress,...) qui assurent l'archivage pérenne des données</p> <p>Pour les données de phénotypes : quels sont les usages ? Plutôt accès via des bases de données dédiées ? Pérennité des données peut être variable</p>					

Commentaires

Ajouter un commentaire à partager avec les collaborateurs

B	<i>I</i>	☰	☰	🔗	📄
Empty text area for comment					

Enregistrer

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 !