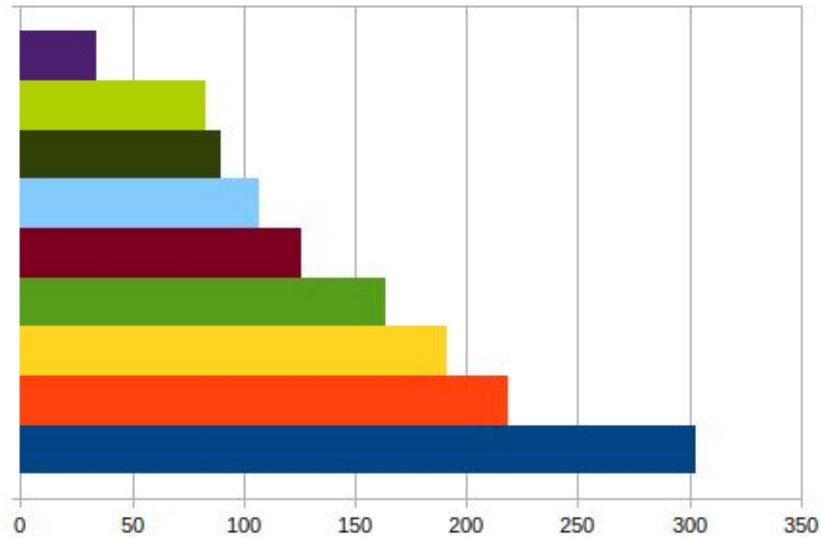


Plan de Gestion des Données

Paulette Lieby

Compétences données/DMP

Compétences recherchées



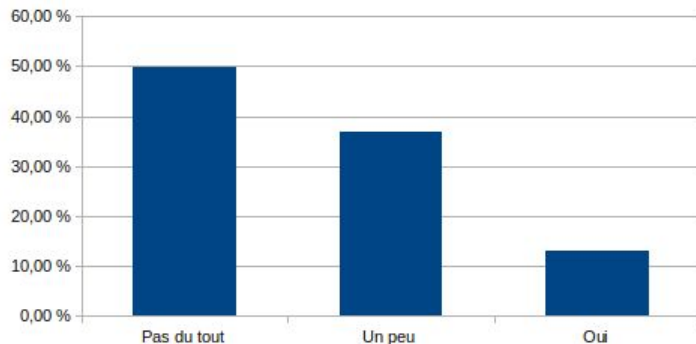
- Analyse bioinformatique / statistique des données
- Développement de workflows d'analyse
- Développement d'algorithmes et/ou d'outils logiciels
- Développement, gestion et administration de bases de données biologiques
- Intelligence artificielle
- Développement de méthodes statistiques
- Implémentation de plans de gestion des données
- Curation de bases de données biologiques
- Autre

Besoin de formation au DMP

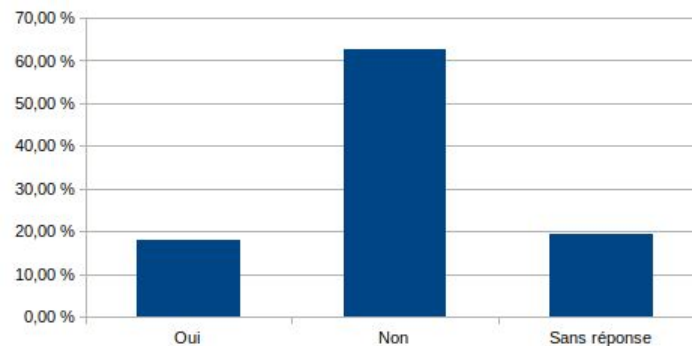


Plan de Gestion de Données (Data Management Plan, DMP)

Etes-vous familier avec le concept de Data Management Plan (DMP) ?



Votre équipe recourt-elle à des DMP pour ses projets ?



Utilisez-vous un modèle de DMP
(principalement Opidor + Modèles INRA)

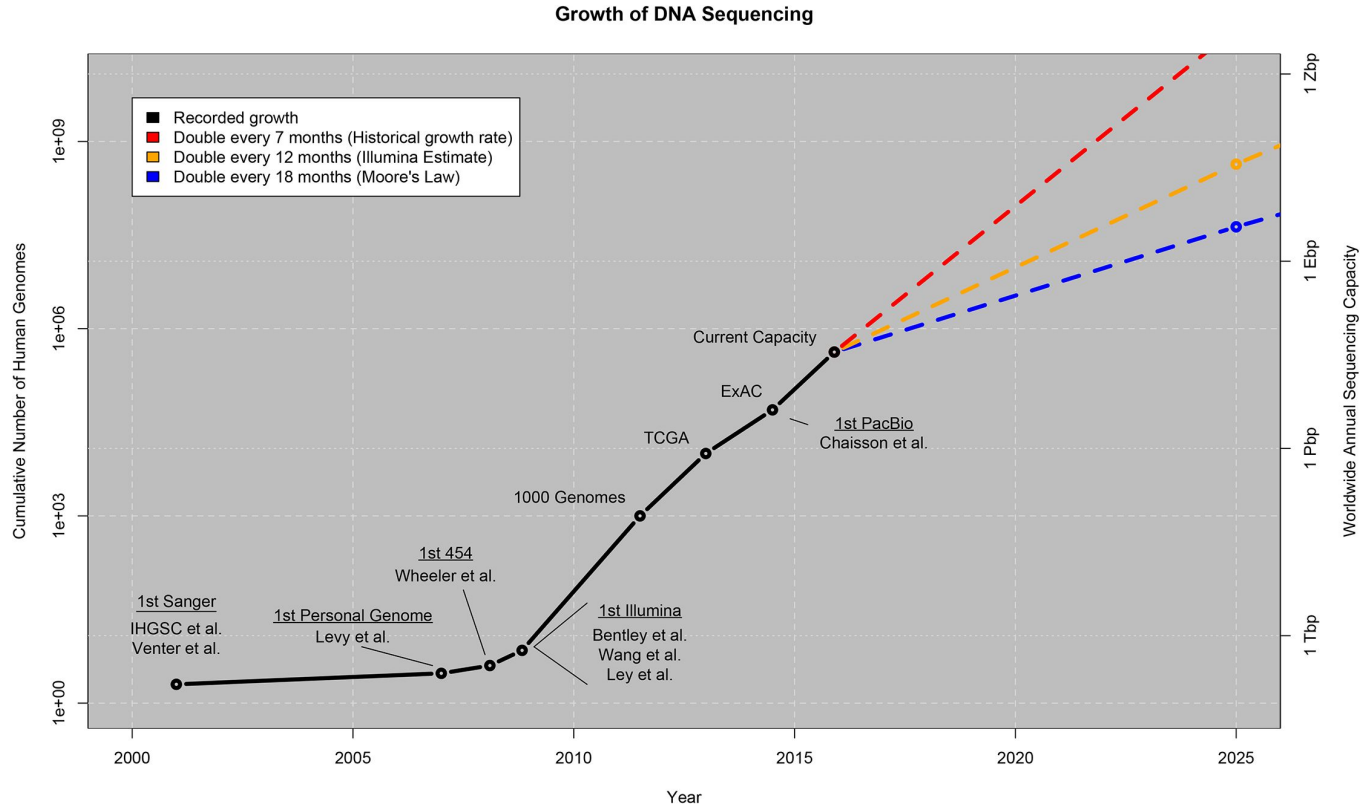


Le contexte

1. Répondre aux besoins exprimés (stockage, gestion des données)
2. L'action 1.6 : "*Partager les services et ressources avec les autres infrastructures de recherche*"
3. La demande de l'ANR d'accompagner chaque projet d'un plan de gestion des données (PGD) à partir de 2019

- ➡ implique que l'IFB s'approprie et déploie la culture DMP
- ➡ rendre nos projets et nos ressources visibles (science ouverte)
- ➡ permettre et faciliter la gestion des données et ressources
- ➡ permettre et faciliter le partage des données et ressources

L'explosion des données en bioinformatique



Zachary D.S. *et al.*, Big Data: Astronomical or Genomical?, 2015

Ooops... où sont les données ?

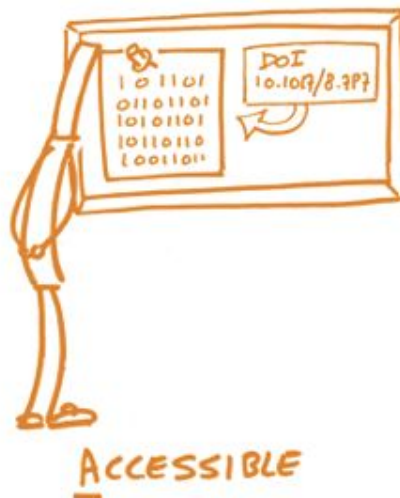


PUBLICATIONS AND DATA

- en croissance exponentielle
 - données complexes
 - structurées ou non-structurées
 - brutes, traitées, ou intermédiaires
 - attributs et formats
 - durée de vie
 - espace de stockage
 - protégées ou non

➡ comment s'assurer que les données soient tracées, plus encore, ouvertes ?

FAIR DATA PRINCIPLES



<https://book.fosteropenscience.eu/>

- Pour que les données soient ouvertes/partageables, il faut qu'elles soient
 - visibles / (re)trouvables : **Findable**
 - accessibles : **Accessible**
 - interopérables : **Interoperable**
 - réutilisables : **Reusable**

➡ **FAIR**

➡ ceci implique une gestion des données depuis la conception du projet et tout au long de son cycle de vie



Australian National Data Service (ANDS)

- un document qui décline les éléments de la gouvernance de données FAIR du projet
- en établissant la transparence du traitement des données, le PGD établit de fait un **contrat de collaboration** entre les acteurs concernés
plateformes ↔ IFB ↔ communautés de recherche
- ce n'est **pas** un autre document administratif à "pondre"
- le document n'introduit pas de contraintes mais permet de décrire et d'organiser le déroulement du projet de recherche
- c'est un document qui **accompagne** le projet tout au long de son cycle de vie
- un document vivant : qui **doit** refléter la réalité du projet (mise à jour)

- easy.DMP développé par EUDAT
 - Data Stewardship Wizard développé par, entre autres, ELIXIR Czech Republic et ELIXIR Netherlands
 - DMPTool US
 - DMPonline UK
 - **OPIDoR FR**
 -
-
- typiquement répondre à une liste de questions pré-établie
 - qui génère ensuite un document PGD/DMP (Data Management Plan)
 - nécessité de disposer d'un modèle de questions construit de manière à ce
 - qu'il permette de décrire les données de manière FAIR
 - qu'il englobe tous les cas de figure pertinents
 - tout en atteignant un niveau de granularité suffisant pour guider efficacement les réponses

- un DMP rédigé en texte libre
 - pas propice à l'automation
 - mise à jour difficile
- ➡ le DMP ne remplit pas sa fonction qui est d'accompagner les données tout au long de leur cycle de vie
- ➡ le DMP ne renseigne aucun des autres acteurs partie prenante dans ce cycle (par ex. stockage) d'où nécessité de duplication de tâches

- DMP automatisé : **mettre à l'œuvre** les informations contenues dans le DMP pour permettre la FAIRification des données
- ➡ un maDMP va permettre au DMP d'être mis à jour et d'être **toujours en phase avec la réalité du projet**
- ➡ **"converting DMPs from a stick to a carrot"** (Simms *et al.* 2018)

- un projet pilote : EMBRC bioimaging
 - rédaction du DMP
- du retour d'expérience, élaborer la structure des éléments du DMP, de manière à ce que
 - le DMP soit générique pour pouvoir être utilisé par tous les acteurs IFB
 - tout en étant précis pour permettre une description complète et détaillée
- le premier domaine sujet à l'automation : le stockage de données (maDMP4LS)
 - ➡ déterminer comment implémenter cette automation (avec OPIDoR)
- formation