



Formation doctorale

—

gestion des données de la recherche

Cellule Data Grenoble Alpes – mai 2022



- **Module principal sur 2 jours :**
 - organisation et description : mardi 3 mai 2022
 - stockage, diffusion et conservation : mardi 10 mai 2022
- **Modules supplémentaires :**
 - gérer des données à caractère personnel : mardi 24 mai
 - diffusion des données : mardi 31 mai
- **Objectifs :**
 - sensibiliser à la gestion FAIR des données de la recherche
 - apporter des recommandations pratiques
 - accompagner à la rédaction du plan de gestion de données
- **Formation interactive :**
 - des présentations classiques, des sondages, des activités
 - **surtout : posez toutes les questions que vous souhaitez !**

Contact Cellule Data : sos-data@univ-grenoble-alpes.fr



Qu'est-ce que la gestion des données de la recherche ?





- Données d'observation
- Données expérimentales
- Données de simulation numérique
- Données dérivées ou compilées
- Données de référence
- Données d'enquêtes
- Code informatique

- **Ensemble des informations collectées, produites et utilisées par la recherche scientifique**
- **Des données très variées dans leur forme**
- **Très dépendantes des disciplines scientifiques**



Ces données (d'observation, expérimentales, de simulation numérique...) ne doivent pas être confondues avec les données au sens du **RGPD**.

Ce dernier distingue, parmi les données, les **données personnelles** et les **données personnelles sensibles**.

Cycle de vie des données de la recherche



Cycle de vie des données, Recherche Data Gouv



- On oublie bien souvent l'étape de **destruction des données**, ainsi que le fait que les données ont un **impact environnemental** !



Le PGD (ou DMP, Data Management Plan) **permet de réfléchir à la gestion des données d'un projet en amont afin de l'anticiper** :

- **Quelles données vont être obtenues** : quels types de données, comment sont-elles collectées, où les stocker, comment on sécurise le stockage, quelle volumétrie, quels formats, quelle organisation, quelle documentation ...
- **Comment vont-elles être utilisées** : comment on les partage, comment on les traite, où on les traite, ...
- **Comment elles vont être préservées** : à quel terme, quelles données, où, comment ...
- **Comment elles vont être valorisées** : comment les diffuser, sous quel format, sous quelle licence, quelles données, comment associer les codes, ...
- **Comment assurer le financement des ressources nécessaires ?**

Pour répondre aux principes FAIR de l'open data



Par défaut, **l'ouverture des données de la recherche est obligatoire selon le principe "aussi ouvert que possible, aussi fermé que nécessaire"**.

Exceptions :

- Données **personnelles** non anonymisées ou pseudonymisées
- Données **non achevées** (sauf données géographiques qui peuvent être inachevées, directive INSPIRE), certaines données environnementales (cf protection des espèces)
- **Photos** quand il y a une personne reconnaissable
- Données **scrapées** (car souvent interdit par les conditions d'utilisation) sauf :
 - si le préjudice du producteur est nul (risque encouru nul également)
 - le scraping d'une partie non substantielle des données reste possible



- En premier lieu **faciliter le travail de recherche** :
 - Assurer l'**exploitation des données** tout au long du projet
 - Assurer la **conservation** des données à moyen terme
 - Pouvoir **reproduire** les résultats scientifiques, y compris par l'équipe qui les a obtenus !
 - Pouvoir facilement les **réutiliser** pour produire de nouvelles recherches
 - **Valoriser** les résultats scientifiques, mais aussi les données et les codes qui ont permis de les obtenir et en augmenter la visibilité
 - Favoriser de **nouvelles collaborations**, de nouvelles approches
- Enjeux d'**ordre sociétaux** :
 - Assurer une **souveraineté** sur les données produites
 - Assurer l'**intégrité scientifique**
 - Garantir la **transparence**, et assurer la confiance des citoyens en la recherche



Tour de table





Présentez vous en quelques mots :

- votre laboratoire et vos thèmes de recherche
- les données manipulées dans votre thèse (type, volume, produite ou réutilisée)
- une attente de la formation
- une question préalable