

Pourquoi rédiger un plan de gestion de données ?

Café Science Ouverte PAMIR

–

16 novembre 2023

Laurent Romary, Inria



Au programme

- Qu'est-ce qu'un plan de gestion de données ?
- Pourquoi un plan de gestion de données ?
- Quelle est la structure d'un plan de gestion de données ?
- Éléments de réflexion

Qu'est-ce qu'un plan de gestion de données ?

Plan de gestion de données (PGD) - *Data Management Plan* (DMP)

- Document introduit (par l'EU à l'origine) pour accroître la prise de conscience du rôle des données dans les projets de recherche
 - Combler le vide en matière de préservation, traçage, réutilisation
- Document décrivant dans les grandes lignes le rôle et la façon dont les jeux de données ont vocation à être créés, décrits, gérés et communiqués au sein d'un projet de recherche
 - Rendu obligatoire par un certain nombre de financeurs : Europe, ANR, DIM PAMIR
 - En général, deux phases : un PGD prospectif en début de projet et une version de synthèse en fin de projet
 - Des environnements dédiés pour la rédaction collaborative des PGD : <https://dmp.opidor.fr>
- Document couvrant les données de recherche, pas les données administratives
 - Niveau grossier de granularité (jeux de données cohérents)
 - Accompagnement des actions à mener sur les jeux de données

Pourquoi un plan de gestion de données ?

- Gérer la complexité des données par rapport aux publications
 - *Aussi ouvert que possible, aussi fermé que nécessaire* : propriété intellectuelle, données personnelles, droit des affaires
 - Variété des jeux de données possibles : complexité, formats, tailles
 - Fragmentation des solutions d'hébergement pour les données
- Accompagner la création d'une culture de la donnée et la mise en place des infrastructures associées
 - Cf. Recherche Data Gouv

Quelle est la structure d'un plan de gestion de données ?

- Description des données
 - Types d'information : cadre de la recherche, réutilisation (sources), création (équipement), enrichissement
- Documentation et métadonnées
 - Sous quelle forme les données vont-elles être décrites ?
- Contraintes portant sur les données
 - Quelle est la sensibilité des données ? Y a-t-il des restrictions à leur diffusion ?
- Responsabilités
 - Qui s'assure de la gestion effective des données ? Quel point de contact pour leur réutilisation ?
- Stockage et conservation
 - Quelles sont les solutions d'hébergement et de diffusion mises en œuvre ou prévues dans le projet ?
- Partage et réutilisation
 - A quelles conditions ces données pourront-elles être réutilisées par des tiers ?

Éléments de réflexion

- Le PGD est un document scientifique
 - Auteur, financeur, publication
 - Un possible préalable à un *data paper* en fin de projet
 - Diffusion (même quand les données ne sont pas diffusables), dépôt dans HAL
- Un PGD est utile, même quand il n'est pas obligatoire
 - Ex. gérer les données dans le cycle court d'une thèse
- Un PGD est un document révisable
 - A mettre à jour quand de nouveaux jeux de données, des perspectives d'hébergement ou de nouvelles contraintes apparaissent au cours du projet
- Et le logiciel ?
 - Intégrer les références aux logiciels utilisés pour créer, transformer, visualiser les jeux de données
- Perspectives
 - Travail plus fin sur l'après ? Cf. Charte de réutilisation des données (DARIAH, DIM PAMIR)
 - Travail en cours sur le PGD actionnable, pour faciliter la réutilisation des informations (métadonnées), le lien avec les publications, etc.