

Comment gérer ses données actives ?

Les données actives sont les données utilisées pour les projets en cours, entre la phase de collecte et de dépôt de données.

Bien gérer ses données actives permet :

- Une meilleure efficacité de l'utilisation des données durant le projet.
- Une réduction de perte de données.
- Une facilité de transfert de dépôt de données en fin de projet.

Cycle de vie des données :



- Planification : phase préalable à la gestion des données, avant même de les avoir en main. Elle consiste à mettre en place un plan de gestion des données (PGD).
- Collecte/création : phase active du projet où l'on rassemble les données brutes.
- Organisation et analyse : phase de nettoyage et de traitement des données, d'analyse et d'interprétation, mais aussi d'organisation des données en dossiers, d'application de convention de nommage, de format, etc.
- Dépôt/Archivage : cette phase consiste à déposer ou archiver les données en fin de projet.
- Publication et partage : une fois que les résultats de recherches ont été publiés, les données peuvent être partagées pour permettre leur réutilisation.
- Réutilisation : les données réintègrent un nouveau projet.

Bien gérer ses données actives en quelques étapes :

1. Identifier les stades de vie des données

Ceci permet de connaître la durée d'utilisation de chaque type de données et de mieux bâtir le PGD.

Ex: distinguer les données brutes, traitées, analysées, finales

2. Hiérarchiser les données selon leur durée de vie

- Cette étape permet de mesurer l'importance de chaque type de données pour mieux prendre les décisions en matière de gestion de données. Ex : les données brutes sensibles peuvent être effacées, une fois qu'elles ont été anonymisées.

3. Se munir de bonnes ressources

- Travailler avec les bureaux de la recherche et d'éthique pour bien préparer les consentements des participants.
- S'assurer de formaliser les ententes de transfert de propriété intellectuelle avec les assistants de recherche.
- Se renseigner sur les types de stockage disponibles (ex : Nearline : bonne capacité de stockage, mais moins rapide; Local scratch : rapide mais pas idéal pour les sauvegardes de longue durée, etc.).



Autres astuces :

- Mettre en place une documentation commune : répertoire des noms et localisation des données et une structure commune pour les données au sein de l'équipe de recherche.
- Créer les métadonnées (données décrivant d'autres données) quand c'est encore frais en mémoire et préciser le logiciel utilisé.
- Instaurer un consentement concernant la réutilisation des données personnelles collectées afin de permettre leur partage.

