

Atelier 2 : Comment gérer des données de recherche dans un monde contraint ?

Préparé et animé par Alexis Arnaud et Violaine Louvet.

Qu'est-ce qu'une gestion durable des données ?

Une gestion durable des données de recherche est une gestion visant à minimiser la production des données de recherche afin de limiter les impacts environnementaux de la recherche, tout en permettant de réutiliser les données.

Pour cela, elle s'appuie sur des pratiques à développer ou renforcer :

- ⑩ Anticiper une gestion durable des données lors des montages de projet, y compris concernant la suppression des données.
 - ↘ Réaliser une analyse en cycle de vie des données, et de la recherche en général, sous l'angle environnemental
 - ↘ Acceptation de la suppression des données, en 1° lieu celles intermédiaires et les doublons.
- ⑩ Réutiliser les données avant d'en produire de nouvelles et favoriser la réutilisation de celles produites :
 - ↘ Inventorier les données existantes pour les réutiliser avant de chercher à en produire d'autres.
 - ↘ Décrire les données de façon détaillée.
 - ↘ Minimiser le poids des formats de données.

La mise en place d'une gestion durable des données de recherche a également besoin :

- ⑩ De métriques et d'indicateurs pour l'évaluation de l'empreinte environnemental des données.
 - ↘ Prise en compte de l'impact du stockage des données, fréquence d'accès, type de support.
 - ↘ Définition de durées de stockage pertinente, notamment en lien avec la finalité
- ⑩ De mutualiser les ressources matérielles et humaines.
- ⑩ De sensibiliser et former les équipes de recherche, en particuliers les doctorantes et doctorants : pour la montée en compétence des équipes de recherche et la création d'une culture de la donnée au sein des laboratoires.
- ⑩ D'une coopération entre les métiers.
- ⑩ De la mise en place de formations pour la montée en compétence des personnels d'accompagnement.
 - ↘ En particulier autour des méthodologies d'analyse en cycle de vie (ACV)
- ⑩ Du développement de politiques institutionnelles et de politiques de laboratoire en matière de développement durable et gestion des données.

Ces problématiques amènent à des réflexions sur la finalité de la recherche elle-même dans un monde très compétitif, mais aussi sur le rôle des financeurs (voire des comités d'éthiques) en particulier pour la mise en place de critères environnementaux dans les appels à projet.

Enfin, un point d'attention concerne la massification des données et la nécessité de devoir passer par des processus automatiques comme l'IA pour pouvoir être en mesure de les exploiter.

Quels sont les freins à une gestion durable des données de la recherche ?

Les principaux freins à une gestion durable des données de la recherche sont avant tout :

- ⑩ L'environnement compétitif de la recherche et du système économique
- ⑩ L'impact des politiques publiques et des financeurs et en particulier :
 - ↘ le financement par projet et globalement la course aux projets et au saut de projet en projet
 - ↘ le manque de moyens pour bien gérer les données en fin de projet et favoriser leur réutilisation
 - ↘ le choix des projets financés (pas de critère environnemental, des priorités autour de thématiques particulièrement impactante comme l'IA ...)
 - ✦ En particulier, il y a une absence de réflexion en ce qui concerne l'utilisation de l'IA dans les projets de recherche et la demande croissante de numérique
 - ↘ la mise en place d'outils ou d'indicateurs mais sans suivi ou action par la suite (par ex. les PGD)
 - ↘ L'absence ou l'invisibilité de certains métiers en lien avec les données (par ex. les archivistes) et le manque de collaboration entre tous les métiers de la donnée.
- ⑩ Les pratiques individuelles notamment :
 - ↘ La course à la renommée, la peur de perdre la paternité de ses travaux
 - ↘ Le conservatisme et la résistance au changement
 - ↘ Les politiques locales, avec potentiellement des conflits de personnalités et du manque de mutualisation et de collaboration.
 - ↘ L'utilisation de matériels ou logiciels propriétaires

Quels sont les leviers pour une gestion durable des données de la recherche ?

Les principaux leviers pour favoriser une gestion durable des données de la recherche concernent :

- ⑩ Les politiques publiques et les financeurs qui peuvent avoir des impacts déterminants :
 - ↘ Rendre les informations indiquées dans les PGD contraignantes, en assurant un suivi et une vérification des propositions formulées
 - ↘ S'assurer de l'avis du comité d'éthique avant chaque projet de recherche pour la mise en balance de l'apport du projet par rapport à ses impacts
 - ↘ Proposer des choix des sujets de recherche financés, notamment favoriser la recherche sur des sujets visant à réduire l'impact environnemental
 - ↘ Proposer des indicateurs pour mesurer et rendre les choses visibles, et permettre aux porteurs de projet d'anticiper les impacts environnementaux
- ⑩ L'évaluation de la recherche :
 - ↘ Avec une prise en compte de cet impact dans les critères d'évaluation des laboratoires et des personnels
 - ↘ Et une valorisation des actions sur la réduction des impacts environnementaux dans la carrière des personnels
- ⑩ Le niveau local où les leviers peuvent concerner :

- ↘ La mise en place de quotas et d'indices carbonés sur les projets
 - ✦ En particulier faire des bilans des données peu ou pas exploitées
- ↘ La montée en compétence des personnels d'appui à la donnée sur ces sujets
- ↘ La sensibilisation et la formation à différents niveaux :
 - ✦ les comités de décision : direction de laboratoire, d'établissement
 - ✦ les jeunes chercheurs et chercheuses
 - ✦ avec des parcours avec des compétences à valider
- ↘ La mutualisation sur différents aspects :
 - ✦ des ressources humaines, des techniques et compétences, ce qui permet aussi d'avoir une montée en expertise
 - ✦ des matériels et infrastructures, avec un point de vigilance sur la gestion des doublons
- ⑩ Au niveau des communautés, il faudrait :
 - ↘ Faire un état de l'art des données existantes

Evolution des PGD pour intégrer une gestion durable des données de la recherche ?

Le PGD peut être un élément important pour avancer dans une gestion plus durable des données de la recherche. Il faut arriver à traduire les éléments précédents pour les intégrer au plan.

Il y a globalement un problème de timing car le PGD est demandé au démarrage du projet et non au montage. De même la version finale est attendue à la fin du projet ce qui ne correspond pas forcément à la publication des articles et des données.