

Table ronde 4 : Les défis de la gestion de données

Les difficultés sont à différents niveaux : à l'acquisition, au stockage et lors de la valorisation des données.

Les points principaux abordés pendant cette table ronde sont : stockage et réseaux ; annotation puis la formation pour l'utilisation de ces outils.

Infrastructures informatiques

Quelles sont les solutions de stockage et qui doit en être responsable ?

- Nous n'avons pas forcément les moyens financiers, humains et en terme d'infrastructure
- Ce n'est pas que notre responsabilité mais aussi celle de la structure (institut, université, hôpitaux . . .)
- Les plateformes doivent accompagner les utilisateurs dans cette prise de conscience mais aussi dans la formation de l'utilisation de certains outils de gestion et stockage de données
- Les instituts et les laboratoires doivent être responsables, le temps ingénieur peut être dédié pour une formation, explication, accompagnement. Cependant le choix, la mise en place et la gestion du système permettant le stockage des données, c'est beaucoup pour un ingénieur de plateforme seul. Cela doit être un travail en collaboration avec les services informatiques.
- Les compétences dépassent notre dimension, notre métier. Nous n'avons pas forcément l'expertise.

Quelle est la capacité des réseaux présents sur notre plateforme ?

- C'est souvent insuffisant.
- Les EPST université et tutelles diverses commencent à en prendre conscience mais ils sont en retard et il y a une urgence
- Il y a une inégalité dans les infrastructures. Les DSI sont-elles suffisamment sollicitées, impliquées, activés et réactives ?
- Il serait intéressant d'intégrer dans le cursus une partie gestion et stockage des big data pour en comprendre l'importance et les rouages.
- OMERO n'est pas utilisé comme serveur de stockage. La virtualisation du stockage = 15To (200 euros / To) pendant 5 ans.
- Mesocentre = hébergement de stockage commun extérieur
- Les tutelles ne sont pas toutes au courant des besoins. Les mésocentres et les solutions locales doivent co-existés.
- Il faut une vraie gouvernance des données qui impliquent tout le monde.

Pourquoi et comment annoter les images :

- Les utilisateurs savent s'organiser et annoter mais pas forcément de la manière que nous voudrions.
- La question est donc de savoir s'il faut forcer les utilisateurs à suivre notre manière d'annoter sinon pas de transfert et de stockage de donnée.
- Ou s'il faut laisser l'utilisateur faire car finalement, la donnée lui appartient et de ce fait, c'est lui qui décide et qui gère. Dans ce cas nous n'avons qu'un rôle de conseil.
- Est-ce que le stockage doit forcément s'accompagner d'annotation ? Dans la plupart des cas, on se rend compte que sans annotations, il devient extrêmement difficile d'exploiter efficacement les données.

- Au départ l'annotation est là pour que l'utilisateur se retrouve.
- L'annotation est là pour faciliter le partage et la compréhension des données
- Toutes données financées par le publique doit être accessible au publique.
- L'annotation et l'accès aux images participent au recyclage des données.
- Les ANR demande que le cycle de vie des data soit évalué et pris en compte lors d'un projet de recherche.
- Les technologies utilisant l'Intelligence Artificielle vont être en demande d'images annotées.

Comment former les utilisateurs et les collègues à ces outils de gestions et de stockage :

- C'est compliqué et c'est long
- Il faut sensibiliser les chefs d'équipe et progressivement former le plus de monde possible.
- Archivage : au bout de combien de temps peut-on effacer les données? C'est un problème émergent. Cela demande une vraie réflexion car toutes ces images sont un vrai puit de connaissance. C'est un patrimoine de savoir.
- Les chefs d'équipe à venir doivent prendre en compte cette dimension pour la gestion et l'utilisation des data
- Il y a un besoin de création de poste pour archiver et gérer les données (des postes existent déjà en astronomie et en science humaines et sociales).