

Proposition de critères de sélection des données à conserver

Cette proposition de critères a été établie à partir de 3 documents :

- NERC : Data Value Checklist - <https://www.ukri.org/publications/nerc-data-value-checklist/>
- DCC : Five steps to decide what data to keep - <https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep>
- Alliance de recherche numérique du Canada : Guide d'évaluation pour la préservation des données de recherche - <https://zenodo.org/records/6283886>

1- Considérations légales, réglementaires, politiques :

- Existe-t-il une raison légale ou juridique de conserver les données ? (les données pourraient-elles être nécessaire en cas de recours juridique ?)
- Faut-il conserver les données pour se conformer à une réglementation ? (informations environnementales par exemple)
- Le financeur ou le bailleur de fonds imposent-ils la conservation des données ? Si oui, sur quelle durée ?
- Existe-t-il une politique de conservation dans votre établissement, votre structure de recherche, votre laboratoire ?

2- Considérations scientifiques et patrimoniales :

- Les données viennent-elles étayer un résultat de recherche publié ?
- Les données sont-elles non reproductibles, difficiles ou coûteuses à reproduire ?
- Les données constituent-elles un jalon dans le progrès des connaissances ? Ont-elles une valeur historique ?
- Les données peuvent-elles constituer un lot de référence pour la communauté scientifique ou disciplinaire ?

3- Contextualisation et documentation des données :

- Le contexte de production ou de recueil des données (comment et pourquoi ont-elles été collectées, comment ont-elles été traitées) est-il suffisamment documenté ? (dans un plan de gestion de données actualisé par exemple)
- Les données sont-elles documentées avec des métadonnées suffisantes et standardisées ?
- Le contenu intellectuel des données est-il renseigné, si possible à l'aide de vocabulaires normalisés ?

4- Contenu et qualité des données :

- Les données ont-elles leur intégrité d'origine ?
- Les données constituent-elles une copie unique ou ont-elles déjà été mises à disposition ou publiées ?

- Les données sont-elles dans des formats ouverts ? S'il s'agit de formats propriétaires, les logiciels ou matériels nécessaires pour utiliser les données sont-ils largement utilisés dans le domaine et facilement disponibles ?
- La qualité des données est-elle suffisante en termes d'exhaustivité, de robustesse, de taille de l'échantillon, de validité, de fiabilité, de représentativité ou de tout autre critère pertinent dans le domaine ?
- Les données contiennent-elles des informations sensibles qui ne doivent pas être diffusées ?
- Les données sont-elles accompagnées de code ou de modèle permettant de les reproduire ?
- Les données sont-elles accompagnées de licence permettant/encadrant leur réutilisation ?

5- Considération budgétaires :

- Disposez-vous de moyens (temps de travail, personnel qualifié) pour valider et/ou améliorer la documentation des données ?
- Disposez-vous de moyens financiers et/ou de solutions matérielles pour assumer les frais de stockage et de conservation des données au-delà de la durée de vie du projet de recherche ?