

Introduction à Curation de données

Qu'est-ce que la curation des données ?

La curation des données¹ est la gestion active des données de recherche lors de leur création, de leur maintenance, de leur utilisation, de leur archivage, de leur partage et de leur réutilisation. Il s'agit d'un processus itératif qui rajoute de la valeur à la recherche en optimisant les ensembles de données pour leur utilisation actuelle ainsi que pour leur découverte et réutilisation futures. Le but majeur de la curation des données est de créer des ensembles de données qui incarnent les principes directeurs FAIR² : Facilement repérable, Accessible, Interopérable et Réutilisable.

Pourquoi faire de la curation de données ?

La curation des données profite à un large éventail de parties prenantes :

LES COMMUNAUTÉS DE RECHERCHE

améliorer la transparence lorsque les données peuvent être correctement validées et répliquées.

FONDS DE RECHERCHE (et le public)

améliorer le retour sur investissement, car les produits de recherches précédentes sont plus facilement découverts et réutilisés dans de nouveaux projets de recherche.

LES CHERCHEUR(EUSE) S INDIVIDUEL(LE)S

faciliter la gestion des données et rendre les collaborations plus faciles, ainsi que bénéficier d'un nombre accru de citations pour les publications et les jeux de données publiés.

OPERATEUR(RICE)S DE DÉPÔTS DE DONNÉES

Bénéficiez de produits de qualité supérieure, plus faciles à découvrir, mieux interconnectés et mieux préparés pour la préservation à long terme.

Qui sont les curateur(rice)s de données ?

Les curateur(rice)s de données ont différents rôles et peuvent interagir avec les données à divers titres tout au long du cycle de recherche :

- Les membres de l'équipe de recherche peuvent se voir attribuer de manière (in)formelle diverses responsabilités de curation des données.
- Les spécialistes en curation de données peuvent travailler au sein d'organisations qui collectent, organisent et partagent des données.
- Les professionnel(le)s de l'information travaillant dans les bibliothèques et autres unités d'appui à la recherche peuvent mener des consultations sur les pratiques et préparer les données à verser dans des dépôts.

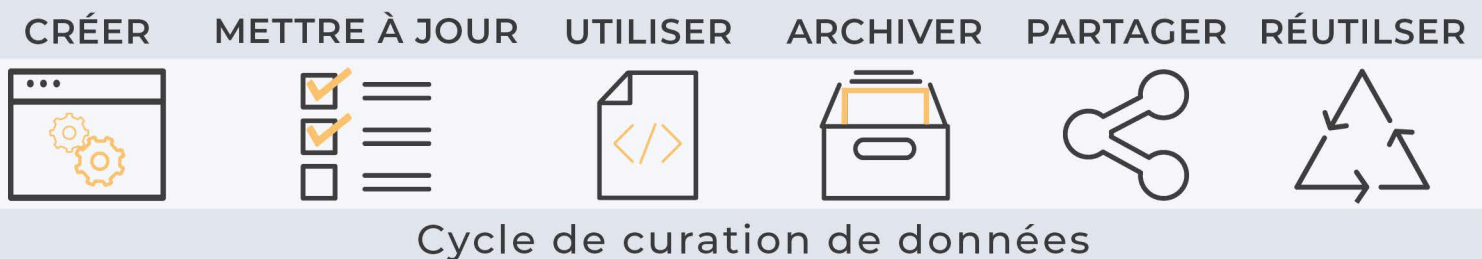
¹ Une définition plus complète en anglais peut être trouvée dans le glossaire du CASRAI, https://dictionary.casrai.org/data_curation

² Voir Wilkinson et al., 2016. doi:10.1038/sdata.2016.18

Que font les curateur(rice)s de données ?

Les curateurs de données appliquent leurs connaissances disciplinaires, leurs compétences en gestion de l'information et leur expertise en logiciels pour améliorer le traitement FAIR des données. Les curateurs de données effectuent un large éventail d'activités³, notamment:

- La préparation à la collecte de données et à la documentation.
- La création de documentation et de métadonnées pour expliquer et contextualiser les données.
- L'implémentation du stockage sécurisé et du contrôle de version pour les fichiers et le code.
- La préparation des ensembles de données en vue de leur versement dans un dépôt de données ou des archives.
- Le contrôle de qualité par la vérification des métadonnées, l'audit de fichier et la révision de code.
- La migration des formats de fichiers pour améliorer la réutilisation et la préservation à long terme.
- L'accroissement des métadonnées et la mise en relation d'ensembles de données pour augmenter la possibilité de découverte.



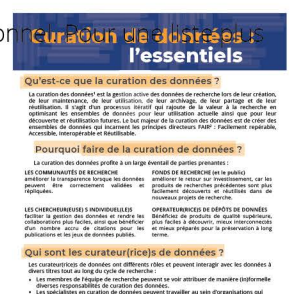
Que fait Portage pour la curation de données ?

Le groupe d'experts sur l'organisation des données⁴ (GEOD) a pour objectif d'élaborer une approche nationale de la curation des données au Canada par le biais de diverses activités, notamment :

- Promouvoir les bonnes pratiques en matière de préparation de données et de métadonnées pour la recherche, la diffusion et la préservation.
- Développer du matériel de formation et mettre en place des ateliers pour améliorer l'expertise en curation de données à travers le Canada.
- Bâtir une communauté de pratique pour les curateur(rice)s de données canadien(ne)s.
- Promouvoir l'adoption de processus standardisés de curation des données dans les dépôts de données canadiens.
- Concevoir un réseau de curation à l'échelle nationale qui exploite les compétences uniques et les connaissances propres à un domaine particulier des curateur(rice)s afin de répondre aux besoins de curation de données de recherche partout au Canada.

³ En référence au modèle établi par le Data Curation Network sur la gestion inter-institutionnelle du personnel, voir : <http://hdl.handle.net/11299/188654>

⁴ Apprenez-en plus sur le GEOD Portage sur <https://portagenetwork.ca/fr>



Le présent guide a été produit par le Groupe d'experts sur l'organisation des données de Portage et peut être modifié et réutilisé en toute liberté en vertu de la licence d'attribution de Creative Commons (CC-BY).

