

STRATÉGIE DE GESTION DES DONNÉES DE RECHERCHE – COLLÈGE LA CITÉ

Contexte

Les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et le Conseil de recherches en sciences humaines (CRSH) sont des organismes fédéraux qui financent la recherche. Ces trois organismes ont publié une politique sur la gestion des données de recherche, car ils sont d'avis que les données recueillies par la recherche au moyen de fonds publics doivent être gérées de manière responsable et sûre. L'objectif de la politique des trois organismes est de promouvoir l'excellence en matière de recherche au Canada en encourageant de saines pratiques de gestion des données de recherche et d'intendance de données (voir section Définitions) dans le respect des obligations éthiques, légales et commerciales en vigueur.

Cette politique est structurée en trois phases :

- **Stratégies institutionnelles** : Les établissements de recherche visés par cette exigence devaient publier leurs stratégies de gestion des données de recherche et en informer l'organisme responsable avant le 1^{er} mars 2023.
- **Plans de gestion des données** : Les organismes subventionnaires devaient déterminer la première série d'occasions de financement assujetties aux exigences du plan de gestion des données avant le printemps 2022. Ils devaient également mettre à l'essai ces exigences dans le cadre d'occasions de financement ciblées avant cette date.
- **Dépôt de données** : Après avoir examiné les stratégies institutionnelles et conformément à l'état de préparation de la communauté de recherche canadienne, les organismes subventionnaires adopteront progressivement l'exigence de dépôt.

La présente directive administrative répond donc à la phase 1, qui est de publier une stratégie institutionnelle.

Importance des données de recherche et de leur gestion

Tel qu'énoncé dans la Politique des trois organismes sur la gestion des données de recherche, celle-ci permet aux équipes de recherche d'organiser, de stocker, de réutiliser les données de recherche numérisées, de même que d'y accéder et d'en tirer profit. Elle est essentielle à la capacité des scientifiques canadiens de conserver et d'utiliser en toute sécurité leurs données de recherche tout au long de leurs projets de recherche, de réutiliser leurs données au cours de leur carrière et, le cas échéant, de les partager. En outre, puisqu'elle est acceptée comme une composante de l'excellence en recherche, l'adoption de solides pratiques en matière de gestion des données de recherche permet aux chercheur.ses. d'assurer une rigueur scientifique et de collaborer dans leurs domaines.

En outre, la gestion efficace des données de recherche est importante, car :

- elle fait partie des pratiques en conduite responsable de la recherche qui est le comportement attendu de quiconque mène des activités de recherche, ce qui inclut la collecte de données (voir PED-15 : <https://www.collegelacite.ca/directives/pedagogique/ped-15>) ;

- elle fait partie également des exigences en éthique de la recherche avec des êtres humaines, telles que décrites dans la PED-13 : <https://www.collegelacite.ca/directives/pedagogique/ped-13> ;
- elle vise à appuyer les chercheur.se.s dans leurs efforts visant à établir et à mettre en œuvre des pratiques de gestion des données conformes aux obligations éthiques, juridiques et commerciales ;
- elle gère leur accès aux personnes autorisées ;
- elle assure leur protection et leur préservation ;
- elle favorise le respect de la confidentialité de certaines données de recherche, telles que celles dans les projets en partenariat avec les entreprises privées ;
- elle assure la conformité du transfert des données entre des membres de l'équipe de recherche ;
- elle permet le partage avec d'autres équipes de recherche, le cas échéant.

Portée

Cette stratégie englobe toutes les données de recherche recueillies sous l'autorité du Collège La Cité, que ce soit dans le cadre des activités d'InnovaCité ou tout autre instance au Collège qui réalise des activités de recherche. Elle s'applique à tout.e chercheur.se, membre du personnel, étudiant.e et collaborateur.trice externe impliqué dans des activités de recherche.

Supervision et révision

Cette stratégie est sous la gouverne du Vice-Président à l'Enseignement et à la Recherche qui délègue sa mise en œuvre à la Direction d'InnovaCité. Elle sera révisée si des changements doivent être apportés et au minimum aux cinq ans.

Soutien institutionnel

Un comité de gestion des données de recherche a été mis en place pour élaborer la présente stratégie, les processus à mettre en place, incluant la gestion des données de recherche. Ce comité est composé de :

- Direction d'InnovaCité
- Deux gestionnaires de centres recherches
 - BIOlab
 - GÉNIElab
- Gestionnaire de la bibliothèque et représentant les technologies de l'information
- Présidence du Comité d'éthique de la recherche avec les êtres humains (CÉR)

InnovaCité sera responsable de sensibiliser les membres du personnel, les étudiant.e.s et les collaborateur.trice.s externe à l'importance de la gestion de données de recherche et assurer l'accès aux outils. Les responsabilités d'InnovaCité incluent :

- Sensibilisation et formation du personnel
- Promotion et soutien des pratiques de GDR
- Accès aux outils, aux ressources et à l'infrastructure de la GDR et soutien de ceux-ci, notamment :
 - Stockage en dépôt pour la consignation des données ;
 - Outils de planification de gestion des données ;
 - Soutien à la curation ;
 - Outils de découverte et d'accès.

Parties prenantes

Le Vice-Président Enseignement et Recherche sera responsable de la diffusion de cette stratégie à l'ensemble du Collège via les canaux de communications existants. Les commentaires pourront être acheminés au Comité de gestion des données de recherche.

Considérations éthiques

Le respect des considérations éthiques seront prises en considération par la présence sur le Comité de gestion des données de recherche :

- La Présidence du CÉR de La Cité pour les questions d'éthiques de la recherche avec les êtres humains ;
- La Direction d'InnovaCité et les gestionnaires des centres de recherche pour les considérations juridiques et commerciales.

Considérations sur les données des autochtones

Les principes PCAP (https://www.afn.ca/uploads/files/nihbforum/info_and_privacy_doc-ocap-fr.pdf) et les principes CARE (<https://mackenziedatastream.ca/fr/article/fair-and-care-data-principles>) seront mis en application pour les projets impliquant de la recherche autochtone.

Autres stratégies et politiques pertinentes

Cette stratégie est liée à la directive administrative PED-15-Conduite responsable en recherche et la PED-13 Éthique de la recherche avec les êtres humains.

Définitions

Les définitions suivantes sont tirées de la Politique des trois organismes sur la gestion des données de recherche: <https://science.gc.ca/site/science/fr/financement-interorganismes-recherche/politiques-lignes-directrices/gestion-donnees-recherche/politique-trois-organismes-gestion-donnees-recherche-foire-aux-questions#1b> et ont été adaptées au besoin.

Dépôt de données : Le dépôt de données se rapporte au moment où les données de recherche recueillies dans le cadre d'un projet de recherche sont transférées à un organe de stockage de données de recherche. Celui-ci devrait contenir des politiques facilement accessibles décrivant les licences de dépôt et d'utilisation, le contrôle de l'accès, les procédures de conservation, les pratiques de stockage et de sauvegarde et les plans de durabilité et de succession. Le dépôt de données de recherche dans les dépôts appropriés appuie la conservation permanente des données dans les limites de l'éthique, des lois et des ententes commerciales et, s'il y a lieu, l'accès aux données.

Idéalement, le dépôt de données comprendra la documentation à l'appui, le code source, le logiciel, les métadonnées et tout autre matériel offrant des renseignements additionnels sur les données et le contexte dans lequel elles ont été recueillies et utilisées pour étayer le projet de recherche. Ces renseignements additionnels simplifient la conservation, la découvrabilité, l'accessibilité et la réutilisation des données.

Données : Les données sont des faits, des mesures, des enregistrements ou des observations recueillies par des chercheurs et d'autres personnes, assortis d'une interprétation minimale de leur contexte. Les données peuvent être présentées sous n'importe quelle forme et sur des supports variés (texte, chiffres, symboles, illustrations, films, vidéos, enregistrements sonores, reproductions picturales, dessins, croquis ou autres représentations graphiques, manuels de procédures, formulaires, schémas, diagrammes, descriptions d'équipement, fichiers de données, algorithmes de traitement, logiciels, langages de programmation, codes et statistiques).

Données de recherche : Les données de recherche sont des données qui sont utilisées en tant que sources à l'appui de travaux de recherche. Les données de recherche peuvent être entre autres des données expérimentales, des données d'observation, des données opérationnelles, des données de tierces parties, des données du secteur public, des données de surveillance, des données traitées ou des données réaffectées. La définition des données de recherche pertinentes est très souvent contextuelle et la détermination de ce qui compte comme tel devrait être guidée par les normes disciplinaires et tenir compte des champs de recherche.

Gestion des données de recherche : La gestion des données de recherche désigne : 1) les processus appliqués tout au long du cycle de vie d'un projet de recherche pour guider la collecte, la documentation, le stockage, le partage et la conservation des données de recherche, et 2) les processus d'utilisation secondaire des données.

La gestion des données de recherche est essentielle tout au long du cycle de vie des données – de la création, du traitement, de l'analyse de la conservation, du stockage et de l'accès, au partage et à la réutilisation (s'il y a lieu), après quoi, le cycle redémarre. La gestion des données devrait être exercée pendant toute la durée de vie des données, notamment la planification de l'enquête, la réalisation de la recherche, la sauvegarde des données au fur et à mesure qu'elles sont créées et utilisées, la diffusion des données et la conservation à long terme des données une fois l'enquête terminée.

Les organismes subventionnaires reconnaissent la diversité des modèles de recherche scientifique et savante qui font progresser les connaissances dans les disciplines représentées par leurs mandats et entre elles. Ils reconnaissent donc qu'il existe des différences légitimes dans les normes de gestion des données de recherche – y compris ce qui est considéré comme des données de recherche pertinentes – entre les disciplines, les domaines de recherche et les modes de recherche qu'ils financent.

Métadonnées : Les métadonnées sont des données sur des données – des données qui définissent et décrivent les caractéristiques d'autres données. Des métadonnées exactes et complètes sont essentielles pour rendre les données de recherche trouvables. Un principe pour aider à déterminer quels renseignements devraient être compris dans les métadonnées est le critère OAIS selon lequel ils devraient être « compréhensibles de façon indépendante ». Autrement dit, suffisamment de renseignements sont fournis dans les métadonnées pour que quelqu'un d'autre soit en mesure de comprendre les données sans que son auteur ait à les expliquer.

Les normes en matière de traitement des métadonnées (souvent appelées « schémas ») sont diverses, mais les métadonnées indiquent généralement qui a créé les données et à quel moment, et varient dans l'ensemble des champs de recherche, et comprennent des renseignements sur la façon dont les données ont été créées, leur qualité, leur exactitude et leur précision, ainsi que d'autres caractéristiques nécessaires pour permettre la découverte, la compréhension et la réutilisation.

Plan de gestion des données : Un plan de gestion des données est un document évolutif – généralement lié à un projet ou à un programme de recherche – qui consiste en des pratiques, des processus et des stratégies rattachés à un ensemble de sujets particuliers touchant la gestion et la conservation des données. Il doit être modifié tout au long d'un projet de recherche pour refléter les changements dans la conception du projet, les méthodes ou d'autres considérations.

Les plans de gestion des données aident les chercheurs à préciser la façon dont ils entendent gérer les données; ils ne les incitent pas forcément à les gérer différemment.

Principes FAIR : Les principes FAIR pour la gestion et l'intendance des données scientifiques constituent une pratique exemplaire internationale pour améliorer la récupération, l'accessibilité, la compatibilité et la réutilisation des biens numériques.

- **Récupérables :** La première étape de la (ré)utilisation des données consiste à les trouver. Les métadonnées et les données doivent être faciles à trouver, tant pour les humains que pour les ordinateurs. Les métadonnées lisibles par machine sont essentielles pour la découverte automatique des ensembles de données et des services.
- **Accessibles :** Une fois que la personne concernée a trouvé les données requises, elle doit savoir comment y accéder, éventuellement par une authentification et une autorisation.
- **Compatibles :** Les données doivent généralement pouvoir être intégrées à d'autres données. En outre, les données doivent être interopérables et capables de fonctionner avec des applications (y compris les logiciels et le matériel informatique) ou des flux de travail pour l'analyse, le stockage et le traitement.

- Réutilisables : L'objectif ultime de FAIR est d'optimiser la réutilisation des données. Pour y parvenir, les métadonnées et les données doivent être bien décrites afin qu'elles puissent être reproduites et/ou combinées dans différents contextes.

Échéanciers

- Publication de la stratégie institutionnelle de gestion des données de recherche d'ici le 1^{er} mars 2023 ;
- Plan de gestion des données de recherche, selon les directives à venir ;
- Dépôt des données de recherche, d'ici 2024.

Regard vers l'avenir

La Cité prévoit développer des processus de sauvegarde des données de recherche qui assurent leur protection et leur confidentialité et étendre ces processus à l'ensemble des activités de recherche.

De plus, La Cité va continuer à se questionner sur l'application de cette stratégie, notamment en ce qui a trait aux données analogiques, aux projets de recherche multi-institutions et autres enjeux qui pourraient être soulevés.

Mise à jour 19 février 2025