

Les données de recherche : un potentiel insoupçonné

Au cours des cinq dernières années, le monde a produit plus de données de recherche que dans toute l'histoire.ⁱ La plupart de ces données sont sous forme numérique, ce qui permet aux chercheurs de leur trouver d'autres fins et des usages novateurs que leurs auteurs n'avaient pas prévus.

Les données de recherche sont les enregistrements factuels servant de sources de première main pour la recherche et qui sont généralement reconnus comme nécessaires par l'ensemble des chercheurs pour valider les conclusions de la recherche.

À l'heure actuelle au Canada, la grande majorité des données de recherche sont perdues parce qu'elles ne sont pas gérées systématiquement. Il existe bien sûr des moyens de gestion des données offerts par des établissements, ou par des organismes nationaux et internationaux pour certaines disciplines et projets de recherche, mais ils ne sont à la disposition que d'une minorité de chercheurs.

En 2004, 34 pays, dont le Canada, ont signé la Déclaration sur l'accès aux données de la recherche financée par des fonds publics de l'OCDE.ⁱⁱ La déclaration repose sur le postulat selon lequel les données de recherche subventionnées par les fonds publics doivent être à la libre disposition du plus grand nombre possible. C'est ainsi qu'un certain nombre de pays cherchent actuellement à déterminer comment ils peuvent exploiter plus systématiquement les données créées par la recherche. Cependant, le Canada n'a pas encore réagi par des mesures concrètes pour donner suite à son engagement avec la déclaration de l'OCDE.

Au Canada, il faut une démarche coordonnée et nationale de gestion des données de recherche pour tirer des avantages plus grands et à plus long terme, à la fois sur le plan social et sur le plan économique, des grands investissements publics qui ont été faits dans la recherche.

Les données sont les matières premières de l'économie du savoir. Leur transformation et retransformation en données de plus grande valeur constituent un bon indicateur de la réussite économique d'une nation.

En 2009, le gouvernement australien a établi l'Australian National Data Service (ANDS) afin d'améliorer la gestion des données de recherche du pays. L'ANDS a pour objectifs d'influencer la politique nationale, de servir de base aux pratiques exemplaires et de transformer les collections disparates de données de recherche en Australie en une collection uniforme de ressources de recherche.

De nombreux organismes savent d'expérience que les coûts de l'entretien des données ne sont rien comparés aux dépenses nécessaires pour remplacer ou recréer des données perdues. Par exemple, les données de l'Inventaire des terres du Canada, inventaire exhaustif et multidisciplinaire des terres du Canada rural, ont été en passe d'être perdues parce qu'elles étaient négligées. Il a fallu quatre ministères fédéraux et plus de cinq ans pour restaurer ces données au prix de milliers de dollars. Il aurait été presque impossible de recréer ces données précieuses qui ont depuis été à l'origine de plusieurs nouvelles initiatives.

Depuis les années 1960, les systèmes d'information géographique (SIG) transforment le processus décisionnel des industries, des gouvernements et des universités en rendant les ensembles de données spatiales et les analyses géographiques numériques accessibles à partir des ordinateurs. Les SIG font partie d'un domaine technologique qui incorpore des données géographiques à des données tabulaires afin de cartographier, d'analyser et d'évaluer les problèmes du monde réel, tels que la planification urbaine, la gestion des biens et l'exploration des ressources. Les SIG sont devenus une industrie immense qui vaut des milliards de dollars.

Les gouvernements des États-Unis et du Royaume-Uni ont récemment lancé des initiatives de données ouvertes qui facilitent l'accès du grand public à des ensembles de données de grande valeur, lisibles par ordinateur, générés par ces administrations publiques. Ces initiatives visent à accroître l'utilisation originale des données générées par le gouvernement à l'extérieur de l'administration publique en encourageant les idées, applications Web et outils novateurs. Elles devraient entraîner des avantages économiques importants, à mesure que les entreprises et les particuliers mettront au point de nouvelles applications et de nouveaux services pour ouvrir de nouvelles voies à l'analyse des données gouvernementales auparavant inaccessibles. Au Royaume-Uni, par exemple, on prévoit dégager une valeur ajoutée de 6 milliards de livres sterling de cette initiative.ⁱⁱⁱ

Avantages d'une bonne gestion et d'un partage des données

Accélérer le progrès scientifique

La bonne gestion des données de recherche permet aux chercheurs d'obtenir et de comprendre les données des autres et de les réutiliser pour leurs propres fins scientifiques, ce qui accélère les nouvelles découvertes.

Accroître la visibilité et les retombées de la recherche

Les données rendues visibles par l'entremise d'un dépôt de données peuvent de beaucoup accroître les retombées de cette recherche.

Respecter les politiques des organismes subventionnaires

De plus en plus d'organismes subventionnaires exigent que les chercheurs et les établissements hôtes conservent, gèrent et rendent accessibles leurs données à la fin d'un projet de recherche.

Éviter la répétition des travaux de recherche

Lorsqu'un ensemble de données est disponible publiquement, il est beaucoup plus probable que ces données ne seront pas créées de nouveau, ce qui évite des travaux coûteux et inutiles de collection et de production des données.

Permettre la reproduction et la validation des résultats de la recherche

Lorsque les données sont archivées et connues, les résultats peuvent être répétés et les données peuvent servir à une nouvelle analyse, ce qui permet de confirmer les premières constatations de la recherche.

L'accessibilité des données de recherche est étroitement liée à l'augmentation du nombre de citations. Par exemple, selon une étude sur le nombre de citations dans les publications sur les essais cliniques sur le cancer, on a constaté que les essais cliniques pour lesquels les données avaient été communiquées ont été cités environ 70 fois plus fréquemment que dans les cas où on ne l'avait pas fait.^{iv}

Intensifier la collaboration

Les données disponibles publiquement permettent aux chercheurs de mieux collaborer les uns avec les autres en mettant en commun les ensembles de données, le contexte et les outils de recherche.

Le contexte actuel

Les méthodes utilisées pour la gestion des données de recherche varient beaucoup selon les disciplines. Dans certains domaines, comme la génomique, la protéomique, la physique des hautes énergies et l'astronomie, des traditions de longue date existent pour l'archivage et la mise en commun des données, ce qui n'est pas le cas dans d'autres domaines tels que la chimie et les sciences humaines et sociales.

L'année polaire internationale (API) constitue le plus grand programme international de recherche scientifique centré sur les régions de l'Arctique et de l'Antarctique jamais connu. Des milliers de chercheurs de plus de 60 nations y ont participé et on a produit des quantités impressionnantes et très diversifiées de données de recherche. Les chercheurs de l'API doivent partager leurs données rapidement et gérer efficacement leurs données de recherche pour faire en sorte que ces données soient accessibles pour d'autres projets de recherche. Le manque de banques de données disponibles pose des difficultés à de nombreux chercheurs de l'API qui essaient de respecter ces engagements.

Les premières archives du genre, telles que la GenBank, ont prouvé le poids que peuvent prendre les anciens ensembles de données pour les nouvelles découvertes, surtout lorsque les données de nombreux laboratoires sont combinées et analysées sous un angle que les premiers chercheurs ne pouvaient pas prévoir.

Victoria Experimental Network Under the Sea (VENUS),^v sous la direction de la University of Victoria, est un observatoire océanique câblé. Il communique des données en temps réel provenant du fond de l'océan par câbles en fibre optique reliés à des instruments à la University of Victoria, en C.-B., où elles sont archivées. Les instruments de VENUS ont envoyé plus d'un demi-milliard de mesures à la University of Victoria et, par l'intermédiaire d'Internet, aux scientifiques et au reste du monde. Pour gérer cette quantité immense de données, la University of Victoria a mis au point un système d'archivage et de gestion des données qui lit et stocke les données et les rend accessibles pendant au moins 25 ans au grand public et aux communautés scientifiques afin de faire progresser la recherche océanique au Canada.

En 2007, le Journal of Applied Developmental Psychology publie un numéro spécial pour souligner l'utilisation, par des chercheurs, des ensembles de données de la Study of Early Child Care and Youth Development pour répondre à plusieurs questions de recherche qui n'avaient pas été prévues dans le premier plan de l'étude. Ces ensembles de données contiennent des données longitudinales sur la famille, la garde de jour et le milieu scolaire de plus de 1 000 enfants, de la naissance à l'âge de 15 ans. La plupart des publications des premiers chercheurs portaient sur des problèmes de garde d'enfants. Cependant, les éditeurs indiquent que les données originales ont été utilisées par des centaines d'autres chercheurs pour répondre à toute une gamme de projets de recherche, tels que le développement de l'enfant, les tendances d'emploi des femmes, la qualité de l'enseignement scolaire, les processus sociaux et psychologiques de la famille, l'interaction entre l'enfant et l'adulte et l'interaction de l'enfant avec ses pairs.

Références

- i. Selon Tony Hey, former Director of the e-Science Core programme in the UK at Digital Preservation Coalition Forum. British Library Conference Centre, London UK, Wednesday 23 June 2004.
- ii. http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html
- iii. http://www.conservatives.com/Policy/Where_we_stand/Technology.aspx
- iv. Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308
- v. <http://www.venus.uvic.ca>
- vi. L'analyse des écarts est disponible au <http://data-donnees.gc.ca/eng/about/achievements.html>
- vii. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers Heather A Piwowar, et. al. PLoS Med. 2008 September; 5(9): e183.
- viii. Charles Humphrey (2004). Preserving Research Data: a time for action. Preservation of Electronic Records: New Knowledge and Decision-making. Ottawa: The Canadian Conservation Institute, pp. 83-90.

La présente brochure a été créée en juin 2010 par Kathleen Shearer au nom du Sous-comité sur la gestion des données de l'Association des bibliothèques de recherche du Canada.

Association des bibliothèques de recherche du Canada

Pavillon Morisset
65, privé Universit , Suite 239
Ottawa Ontario Canada
K1N 9A5
T l phone: 613.562.5385
T l copieur: 613.562.5297
Courriel: carladm@uottawa.ca
www.carl-abrc.ca



  2010 ABRC. Le contenu dans ce document est sujet   une licence de Creative Commons Paternit  - Pas d'Utilisation Commerciale 2.5 Canada.

Lacunes dans la gérance des données au Canada

Le Groupe de travail sur la stratégie des données de recherche dirigé par l'ICIST (Institut canadien de l'information scientifique et technique) a fait à l'automne de 2008 une analyse des écarts dans la gérance des données de recherche au Canada.^{vi} Les obstacles relatés étaient les suivants :

Politiques sur les données: Les politiques sur les données au Canada ne portent pas sur tous les types de données de recherche et, dans certains cas, des chercheurs ne se conforment pas aux politiques.

Financement de la gestion des données: La structure de financement de la recherche au Canada ne prévoit pas la gestion des données au-delà de la durée de vie d'un projet de recherche.

Dépôts de données: Il existe seulement quelques dépôts de données actifs au Canada permettant aux chercheurs d'y déposer leurs données.

Compétences: Les chercheurs ont rarement les compétences nécessaires pour bien gérer leurs données et il y a peu des spécialistes en gestion des données pour les assister.

Normes: Ce ne sont pas tous les chercheurs ni toutes les disciplines qui ont adopté les normes internationales.

Récompenses et reconnaissance: Il y a peu de mesures incitatives favorisant le partage des données par les chercheurs.

Rôles et responsabilités: À l'exception de certains ministères, il n'y a pas d'organisme national chargé de la conservation, de la gestion et de la diffusion des données de recherche.

Temps requis: Il faut une description des données et une documentation qui soient conviviales, claires et détaillées, mais aussi complètes. Ceci requiert du temps et la plupart des chercheurs n'en ont pas beaucoup.

Le manque d'incitatifs pour les chercheurs est considéré comme un obstacle crucial et non résolu à l'établissement d'une culture de partage des données. Afin d'y remédier, la communauté de la recherche doit mettre au point des mesures de partage de gestion des données et procéder à leur suivi. Les contributions au partage de données pourraient ensuite être prises en compte par les établissements au moment de prendre des décisions sur l'embauche, la période d'affectation et la promotion des chercheurs.^{vii}

Politiques pertinentes sur les données de recherche au Canada

Instituts de recherche en santé du Canada:

Exigent des bénéficiaires de subventions qu'ils déposent certains types de données – données en matière de bioinformatique, d'atomistique et données de coordonnées moléculaires – dans les bases de données publiques correspondantes dès la publication des résultats de leur recherche. Les IRSC exigent aussi des chercheurs qu'ils conservent les ensembles de données originales découlant de la recherche qu'ils subventionnent pendant au moins cinq ans après la fin de la subvention.

Conseil de recherches en sciences humaines:

A une politique d'archivage des données de recherche depuis 1990. Selon cette politique, « Toutes les données de recherche rassemblées grâce au financement du CRSH doivent être conservées et mises à la disposition des autres chercheurs dans des délais raisonnables. » Selon le CRSH, un délai raisonnable est une période de deux ans après la fin du projet de recherche pour lesquels les données ont été recueillies. Le CRSH n'a pas appliqué cette politique auprès des chercheurs.

Énoncé de politique des trois conseils sur l'éthique de la recherche avec des êtres humains:

« Les chercheurs devraient s'assurer que les données obtenues sont conservées avec toutes les précautions nécessaires particulières dues à la nature délicate des renseignements. Les données publiées ne devraient contenir ni nom, ni initiales, ni aucune autre sorte de renseignement pouvant mener à une identification. Il peut s'avérer important de conserver certains types d'identificateurs (p. ex., région de résidence), mais ceux-ci devraient être dissimulés le mieux possible, selon un protocole normalisé, avant que les données ne soient communiquées aux fins de la recherche. »

Conseil de recherches en sciences naturelles et en génie:

N'a pas de politique générale relativement aux données de recherches. Il y a cependant des lignes directrices pour les chercheurs subventionnés par l'intermédiaire du Programme de subventions de réseaux stratégiques: « Pour encourager le partage des données et leur utilisation par d'autres, on doit conclure dès le début une entente établissant la responsabilité en matière de mise à jour et de conservation de grands ensembles de données. »

Malgré la politique du CRSH qui prévoit la préservation des données, les études sur les projets de recherche financés par ce même organisme concluent que moins de 3 % des projets avaient archivé leurs données dans une banque, et que ces banques étaient toutes situées aux États-Unis.^{viii}

Exemples d'autres politiques sur les données

Pêches et Océans Canada

« Les gestionnaires de Pêches et Océans ont la responsabilité de s'assurer que les collecteurs de données sous leur direction transmettent leurs données ainsi que les données recueillies en vertu d'un contrat ou de partenariat avec d'autres organismes au centre de données appropriées de façon opportune. »

Année polaire internationale

La Politique sur les données de l'API indique qu'en vue de maximiser les bénéfices des données recueillies sous l'égide de l'API, le Comité conjoint de l'API exige que ces données, y compris les données opérationnelles communiquées en temps réel, soient accessibles intégralement, gratuitement, librement et dans le plus bref délai possible.

Réponses aux préoccupations des chercheurs et des administrateurs

Je ne peux pas archiver et communiquer mes données parce qu'elles portent sur des humains.

L'Énoncé de politique des trois conseils sur l'éthique de la recherche avec des êtres humains précise que « D'une façon générale, la meilleure façon de protéger la confidentialité des renseignements personnels passe par l'anonymat. Si les données conservées sont véritablement anonymes, la recherche ne fera l'objet que d'une évaluation minimale du CÉR [Comité d'éthique de la recherche]. »

Si je partage mes données, je limite mes options d'exploitation de brevets et d'autres avantages économiques.

Les bons principes de gestion des données ne sont pas incompatibles avec ceux qui régissent la propriété intellectuelle et la propriété des données. Les chercheurs peuvent choisir de restreindre l'accès à leurs données afin de maximiser leurs avantages professionnels et économiques en reportant la communication des données jusqu'à la publication ou à l'application du brevet, ou en faisant appel à une licence d'utilisation non commerciale pour les données.

Les bonnes pratiques de gestion des données sont trop coûteuses.

Le maintien d'un cadre fiable et ordonné pour protéger les investissements considérables consentis par les établissements dans la création de données de recherche représentent un coût relativement modeste comparativement à celui prohibitif de recréer ces données.

Les chercheurs dans certains domaines sont réticents à rendre accessibles leurs données pour les raisons suivantes : inquiétudes au sujet de la propriété des données ou de la propriété intellectuelle; méconnaissance de la valeur ajoutée que représente le partage des données ou manque de connaissance ou de temps pour préparer les données à des fins de diffusion.

Que peut-on faire sur le campus?

Chercheurs

- ° S'engager à communiquer les données de recherche le plus ouvertement possible.
- ° Élaborer un plan de gestion des données avant d'entamer un projet de recherche.
- ° Comprendre les politiques de gestion des données des organismes subventionnaires et s'y conformer.
- ° Faire appel aux spécialistes des données sur le campus pour de l'aide dans la collecte et la gestion des données de recherche.
- ° Repérer et utiliser les normes internationales pour la gestion des données.

Administrateurs universitaires

- ° Élaborer des politiques qui favorisent la bonne gestion des données.
- ° Appuyer la mise en place de dépôts de données par l'établissement.
- ° Donner une formation aux chercheurs au sujet des pratiques de gestion des données.
- ° Aider les chercheurs en embauchant des spécialistes ou des bibliothécaires experts en matière de données et faire connaître ces spécialistes aux équipes de recherche (et les y intégrer).
- ° Reconnaître l'apport du partage des données dans l'embauche, les promotions et les nominations.

Bibliothèques de recherche

- ° Mettre sur pied et gérer des dépôts de données pour l'établissement.
- ° Favoriser la formation des bibliothécaires dans le domaine de la gérance des données.
- ° Venir en aide aux chercheurs en embauchant des bibliothécaires experts en matière de données et en mettant ces spécialistes à la disposition de l'ensemble des chercheurs.
- ° Assurer la formation des chercheurs au sujet des pratiques de gestion des données.



Les données de recherche : un potentiel insoupçonné

Une brochure de sensibilisation réalisée par l'Association des bibliothèques de recherche du Canada (ABRC)