

Gérance des données de recherche au Canada : une analyse des écarts

Octobre 2008

Groupe de travail sur la stratégie des données de recherche
<http://data-donnees.gc.ca/fra/index.html>

Introduction

La nécessité d'établir une approche coordonnée de la gérance des données de recherche au Canada a été bien documentée dans une série de rapports publiés au cours des dix dernières années. Pendant cette même période, des milliards de dollars ont été investis dans la recherche au Canada, ce qui a généré d'énormes volumes de données numériques. Collectivement, si elles sont bien réutilisées et si elles sont gérées de façon appropriée, ces données représentent un atout important et sont des sources quasi illimitées de nouvelles connaissances. À ce jour, le Canada n'a pris aucune mesure pour instituer, de façon coordonnée, des pratiques et des services dédiés à la gérance des données de recherche. Par conséquent, des données précieuses sont sous-utilisées et risquent d'être perdues.

En janvier 2008, un groupe de travail composé de représentants de divers organismes canadiens de recherche a été formé avec pour mandat de fournir des recommandations et un plan d'action pour une approche nationale de la gérance des données de recherche au Canada. Le groupe de travail a convenu que la meilleure façon de procéder était de former plusieurs groupes d'étude qui doivent produire un bref énoncé du problème, élaborer une stratégie pour l'éducation et la formation des chercheurs, et réaliser une analyse des écarts ou des lacunes des activités de gérance au Canada. Le présent rapport fournit les résultats de l'analyse des écarts réalisée au printemps-été 2008.

Objectif et méthodologie

L'analyse des écarts a pour objectif de déterminer les écarts entre l'état actuel et l'état idéal. Les résultats serviront à justifier l'invitation à passer à l'action et contribueront à l'élaboration d'une stratégie pratique pour l'amélioration de la gérance des données au Canada. Il ne s'agit pas de reproduire ce qui a déjà été fait, mais plutôt de rassembler tous les renseignements et les preuves existants afin de décrire l'état actuel de la gérance des données de recherche au Canada. Ensuite, une stratégie sera élaborée pour les prochaines étapes. Ainsi, une bonne partie des renseignements contenus dans le présent document proviennent de rapports précédents traitant des données de recherche. Plusieurs publications canadiennes, toutes fondées sur des consultations poussées auprès des intervenants, viennent étayer cette enquête : le rapport final de la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS); le rapport de la Consultation sur les archives nationales de données (CAND); la Stratégie canadienne sur l'information numérique (SCIN). Un certain nombre d'autres rapports ont également été consultés, y compris le rapport de l'Association des bibliothèques de recherche intitulé *To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering*, le document de la U.S. National Science Foundation

intitulé *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, le document de l'UKOLN intitulé *Dealing with Data: Roles, Rights, Responsibilities and Relationships*, le projet de rapport de l'Association des bibliothèques de recherche du Canada intitulé *Enquête sur les initiatives canadiennes et internationales en matière de la gestion des données*. De plus, des renseignements provenant d'autres sources (p. ex. sites Web) et les commentaires des membres du groupe de travail ont été recueillis.

Cycle de vie des données

Le modèle du cycle de vie fournit un cadre utile pour l'analyse de l'état de la gestion des données au Canada. Le cycle de vie illustre les diverses étapes par lesquelles passent les données à mesure qu'elles sont recueillies, utilisées, préservées et réutilisées, assurant ainsi la maximisation de leur valeur. Aux fins du présent rapport, un modèle simple de cycle de vie en quatre (4) étapes a été retenu : production des données; diffusion des données; gestion à long terme des données; découverte et recyclage des données.

2. DIFFUSION

3. GESTION À LONG TERME

1. PRODUCTION



4. DÉCOUVERTE ET RECYCLAGE

Chaque étape comprend un groupe de processus connexes dans le cycle de vie des données. Dans le cadre d'une étape, certains processus ou activités, vus collectivement, constituent un élément important de la conduite de la recherche. Tandis que certaines activités et certains produits sont intrinsèques à chaque étape, d'autres évoluent à travers celles-ci. Par exemple, la conception d'une expérience ou d'une enquête fera partie intégrante de l'étape de la production des données, tandis que les produits de données émanant de cette étape évolueront à travers tout le modèle. Il est à noter que ce modèle vise à faciliter la détermination des écarts dans les activités liées aux données, plutôt qu'à fournir un compte rendu complet et détaillé de la gestion des données dans toute sa complexité.

1. Production des données : Cette étape comprend toutes les activités liées à la planification, à la collecte, au traitement, à l'analyse et à la maintenance des données dans le cadre du projet de recherche initial. Parmi ces activités figure la sélection d'une conception d'étude; l'élaboration d'outils pour la collecte des données; la collecte ou la création des données; la préparation, la vérification et la validation des données; l'analyse des données; la sauvegarde des versions de données; la préparation et le balisage des métadonnées.
2. Diffusion des données : Cette étape comprend la préparation des données pour l'utilisation par autrui et l'établissement de procédures et de méthodes pour la diffusion des données. Afin d'être compréhensibles, les données doivent être accompagnées de métadonnées descriptives accessibles dans des formats répandus. Les données doivent également être disponibles dans des formats communément utilisés, et des mécanismes doivent être en place pour assurer les niveaux appropriés d'accès, selon les restrictions relatives à la sécurité, à la protection des renseignements personnels ou à la propriété intellectuelle.
3. Gestion à long terme des données : Cette étape comprend les activités de soutien à la conservation qui assurent l'accès à long terme aux données de recherche. Le travail de développement des collections de données est essentiel à cette étape et fait partie intégrante de la gestion de l'accès aux données de recherche. L'accès à long terme ne sera possible qu'avec le concours de saines pratiques de conservation. Parmi ces activités figurent l'évaluation, la sélection, le dépôt ou l'absorption des données dans un dépôt, l'assurance de l'authenticité, la gestion de la collecte des données et des métadonnées, l'entretien des médias numériques et la migration des données vers de nouveaux médias numériques.
4. Découverte et recyclage des données : On reconnaît de plus en plus les données comme un atout précieux pour la recherche dont la valeur dépasse l'objectif visé initialement. Cette étape comprend des mécanismes et des activités qui permettent la découverte et la réutilisation des données. Qu'il s'agisse de répéter des constatations antérieures ou de traiter des questions de recherche non explorées, les outils de découverte sont nécessaires pour repérer et récupérer les données pertinentes. Le recyclage des données, qui comprend la création de nouvelles données en combinant les données appropriées provenant d'une variété de fichiers existants, est en voie d'émerger comme nouvelle méthode. Le recyclage des données génère également de nouveaux produits de données qui n'existaient pas auparavant. Parmi les activités de cette étape figurent l'élaboration et le soutien des outils de recherche qui utilisent des métadonnées normalisées, l'harmonisation du codage des données pour des variables précises, la création de nouvelles méthodes de combinaison des données, ainsi que la génération et la récolte de nouvelles collections de données.

Analyse des écarts

La liste d'indicateurs suivante a été utilisée pour réaliser l'analyse des écarts dans la gérance des données au Canada. Dans le cadre de chaque indicateur, l'analyse fournit un bref énoncé de l'état idéal de la gérance des données, une description de la situation actuelle et une discussion concernant les écarts.

- I. Politiques
- II. Financement
- III. Rôles et responsabilités
- IV. Dépôts de données [numériques fiables]
- V. Normes
- VI. Compétences et formation
- VII. Systèmes de récompense et de reconnaissance
- VIII. Recherche et développement
- IX. Accessibilité
- X. Conservation

Les états idéaux, sciemment élevés, sont fondés sur les intrants provenant du groupe de travail. L'analyse ne traite pas des détails précis d'un état idéal puisque ce travail sera l'objet de la prochaine étape du projet. Les états actuels ont été déterminés à partir d'un examen de la documentation existante, ainsi que des intrants provenant du groupe de travail. Une fois les écarts entre les deux états déterminés, un « niveau d'écart » leur a été attribué afin d'indiquer l'ampleur de cet écart.

Dans un certain nombre de cas, il a été impossible de fournir un compte rendu complet de l'état actuel en raison de l'absence d'information disponible. Par conséquent, une certaine variation existe dans l'intégralité des états actuels et des écarts dont il est question dans ce rapport. Cette absence d'information, qui constitue un écart important en soi, a été mentionnée dans les sections pertinentes.

I. Politiques sur les données

État idéal : *Les organismes canadiens disposent de politiques cohérentes fondées sur des principes de gestion des données sains, qui régissent la gestion des données à l'échelle des disciplines tout au long du cycle de vie de la gérance des données et qui reflètent les initiatives législatives pertinentes.*

État actuel

La liste des politiques sur les données de recherche examinées aux fins de ce rapport est jointe en annexe. En résumé :

- aucune politique de gestion de données obligatoire n'existe au Canada;
- bon nombre d'organismes participants à la recherche ne disposent actuellement pas de politiques sur la gestion des données de recherche;
- souvent, lorsque de telles politiques existent, les principes sur lesquels elles reposent ne sont pas explicitement énoncés;

- dans certains cas, les politiques des conseils de l'éthique et les politiques relatives à la protection des renseignements personnels, qui exigent la destruction des données, contredisent les politiques des universités et des organismes de financement, qui exigent la conservation et le partage des données;
- les politiques d'archivage des données de certains organismes de financement (p. ex. Conseil de recherches en sciences humaines du Canada [CRSH]) ne sont pas exécutoires, parce que les organismes n'ont pas mis en place des mécanismes permettant de documenter la conformité aux politiques;
- la majorité des politiques existantes prescrivent seulement la période durant laquelle les chercheurs doivent conserver leurs données, et être prêts à les partager, après l'achèvement d'un projet de recherche.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Dans un nombre limité de cas (p. ex. API*)	Oui, mais parfois en contradiction les uns avec les autres – universités, conseils de l'éthique, <i>Loi sur la protection des renseignements personnels</i> , LPRPDE*	Dans un nombre limité de cas – IRSC*, CRSH*, CRSNG*, organismes fédéraux et provinciaux, certains projets de recherche (API*)	Dans un nombre limité de cas – CRSH*, IRSC*, organismes fédéraux et provinciaux

*Voir le glossaire.

Écarts

Il existe d'importants écarts dans les politiques à l'échelle du cycle de vie des données. Peu de politiques abordent la nécessité pour les chercheurs d'élaborer des plans de gestion des données. En ce qui concerne la diffusion et la gestion à long terme des données, bien que les trois Conseils disposent de politiques de partage des données, elles ne précisent pas que les chercheurs devraient adopter des normes spécifiques afin de garantir l'accès aux données et leur conservation à l'avenir. Seuls quelques organismes disposent de politiques sur l'enregistrement des données dans des dépôts.

Niveau d'écart Modéré

II. Financement

État idéal : *Pris dans son ensemble, l'éventail des mécanismes de financement couvre les coûts tout au long du cycle de vie des données, ce qui assure un appui à long terme pour les nombreux intervenants de la recherche.*

État actuel

Le financement de la gérance des données au Canada provient d'une variété de sources et vise différentes étapes du cycle de vie. La liste des sources de financement de la gérance des données examinées aux fins de ce rapport est jointe en annexe. En résumé :

- Par l'entremise de subventions à la découverte, les trois Conseils fournissent des fonds aux équipes de recherche pour la production et la maintenance des données pendant la durée du projet et, dans certains cas, la diffusion des données;
- Le CRSH offre des fonds pour la gestion à long terme des données, mais peu de subventions ont été approuvées à cette fin;
- La FCI fournit des fonds pour les dépenses en immobilisations des bases de données et l'acquisition des données et, par l'entremise du fond de gestion des infrastructures (IOF), verse certaines sommes pour la maintenance des données pendant la durée d'un projet;
- La FCI finance le Programme des centres de données de recherche (CDR) pour la découverte et le recyclage des données; de plus, le CRSH et les IRSC fournissent aux CDR des fonds d'exploitation s'élevant à un maximum de 50 % de leurs dépenses d'exploitation; les universités fournissent le reste des fonds; CANARIE (par l'entremise de la plateforme sur réseau) fournit certains fonds pour les outils de découverte et de recyclage;
- Le financement de la gestion à long terme des données est soutenu par l'intermédiaire de la gestion des dépôts de données par les ministères gouvernementaux, les bibliothèques de données des universités et les organismes de bienfaisance pour la recherche, ainsi que par les organismes de financement des trois Conseils, mais ces dépôts acceptent uniquement certaines collections de données dans certaines disciplines.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Oui – IRSC, CRSNG, CRSH, FCI, d'autres sources de financement de la recherche	Dans un nombre limité de cas – IRSC, CRSNG, CRSH, FCI	Dans un nombre limité de cas – universités, instituts de recherche sans but lucratif, organismes gouvernementaux, secteur privé	Dans un nombre limité de cas – FCI, CANARIE, CRSH, IRSC, un certain soutien provenant des universités pour l'infrastructure

Écarts

En ce qui concerne le financement des diverses étapes du cycle de vie des données, il existe d'importants écarts dans les étapes de diffusion, de gestion à long terme et de découverte et recyclage des données. Plus précisément, les coûts liés à la préparation de la diffusion des données ne sont pas appuyés par les mécanismes de financement existants et peu d'institutions fournissent un financement soutenu pour les dépôts de données.

Niveau d'écart

Important

III. Rôles et responsabilités

État idéal : *Les rôles et les responsabilités sont clairement définis et entièrement assumés. Chaque participant au cycle de vie des données assume un ensemble distinct de responsabilités et doit agir de concert avec d'autres participants dans la poursuite des objectifs globaux et importants pour l'ensemble de la collectivité en matière de gérance des données.*

État actuel

L'annexe renferme une liste des rôles et responsabilités définis aux fins de ce rapport. En résumé :

- les responsabilités majeures sont assumées par les chercheurs principaux, qui doivent fréquemment (en vertu des politiques des organismes de financement et des universités) s'assurer que les données de recherche sont conservées et partageables durant une période définie (de 2 à 5 ans) après l'achèvement d'un projet de recherche;
- certains organismes gouvernementaux sont responsables de la collecte et de la conservation de certains types de données : Statistique Canada, Ressources naturelles Canada, Agence spatiale canadienne, Environnement Canada, etc.;
- dans un nombre limité de cas, des institutions de recherche (universités, autres centres de recherche) ont assumé la responsabilité de la gestion à long terme des données de recherche.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Équipe de recherche	Équipe de recherche	Certains organismes gouvernementaux, universités, milieux de recherche	Certains organismes gouvernementaux, universités, milieux de recherche

Écarts

Il n'existe pas d'axes de responsabilité en matière de conservation dans le cadre de chacun des éléments du cycle de vie des données, ni de mécanismes de rassemblement des intervenants afin de définir les rôles et les responsabilités. Les organismes de financement de la recherche au Canada exigent rarement l'élaboration de plans de gestion des données qui décrivent les mécanismes et les calendriers selon lesquels les données générées par la recherche seront rendues disponibles. À l'exception de certains organismes

gouvernementaux, aucune institution nationale n'est chargée de la conservation, de la gestion et de l'accessibilité publique des données de recherche dans la mesure nécessaire pour appuyer les besoins des groupes d'intervenants.

Niveau d'écart

Important

IV. Dépôts de données [numériques fiables]

État idéal : *Le Canada dispose d'un réseau complet de dépôts de données numériques fiables qui offre un accès continu et fiable à toutes les données de recherche considérées comme ayant une valeur durable.*

État actuel

L'annexe renferme une courte liste des dépôts de données au Canada qui ont été examinés aux fins de ce rapport. En résumé :

- les dépôts existants au Canada sont gérés par les organismes fédéraux et provinciaux, les milieux de recherche, les universités et le secteur privé;
- il existe un réseau, très récemment établi, de dépôts institutionnels qui sont gérés par les bibliothèques de recherche, mais seuls quelques-uns peuvent recueillir les données de recherche d'une manière qui en facilite l'accès et la réutilisation;
- parmi les dépôts existants, peu ou aucun ne s'inscrit dans la définition du Dépôt numérique fiable (DNF), qui prévoit une politique, un processus, des normes et un cadre technologique pour la conservation numérique;
- de plus, il n'existe aucun processus d'accréditation pour établir la fiabilité d'un dépôt de données et il est possible que des dépôts de données ne puissent pas conserver les données numériques même si leurs caractéristiques portent à croire qu'ils le peuvent.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Dépôts à court terme conçus pour appuyer l'analyse des données	Dans un nombre limité de cas (p. ex. IDD*)	Absence de couverture, mise en œuvre limitée d'activités saines liées à la conservation des données	Dans un nombre limité de cas

*Voir le glossaire.

Écarts

Il existe d'importants écarts dans la couverture et la capacité des dépôts de données. La plupart des ensembles de données de recherche ne sont jamais versés dans un dépôt de données. Il n'existe pas de dépôts de données pour tous les domaines et la vaste majorité des données de recherche sont toujours

conservées sur les disques durs des ordinateurs des chercheurs. Seuls quelques dépôts de données actifs au Canada permettent aux chercheurs d’y consigner leurs données.

Niveau d’écart

Important

V. Normes

État idéal : *Des normes (objet, processus et instrumentation) sont adoptées de façon généralisée dans le cadre de chacune des étapes du cycle de vie des données et elles sont mises en œuvre indépendamment du domaine.*

État actuel

Certains détails relatifs à l’utilisation de normes dans diverses disciplines ne sont pas facilement disponibles. Toutefois, la documentation indique que :

- il existe divers niveaux d’adoption des normes en matière de données au Canada; les centres de données gouvernementaux, universitaires et communautaires importants tendent à adopter des normes axées sur les disciplines;
- pour les collections de données de moindre envergure, il semble que les niveaux de conformité aux normes soient peu élevés;
- il est pratique courante pour les chercheurs d’adopter les critères de production établis par les programmes logiciels, plutôt que de se conformer aux normes internationales;
- dans le domaine des sciences sociales, les centres de données utilisent régulièrement les normes de métadonnées de l’Initiative de documentation des données (IDD), mais aucun modèle général pour la représentation des métadonnées d’études scientifiques n’a émergé au Canada.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Normes axées sur les disciplines	Variable – IDD dans le domaine des sciences sociales; information inconnue pour les autres domaines de recherche	Variable – utilisation régulière dans les grands centres de données (p. ex. archives des milieux de recherche, des universités ou des gouvernements) et mise en œuvre ponctuelle dans d’autres cas	Langage XML souvent utilisé dans le domaine des sciences sociales

Écarts

Au Canada, aucun organisme national ne surveille, supervise ni n’approuve des normes précises qu’utilisent les chercheurs canadiens. Il n’existe aucune stratégie coordonnée de participation à l’élaboration de normes internationales de données de recherche dans le cas des projets de métadonnées, tels que l’Initiative de documentation des données, ni d’outils pour accéder aux données, tels que le projet *Networked Social Science Tools and Resources* (NESSTAR), et de projets collectifs d’infrastructures internationales, comme Frameworks de l’Union européenne. Beaucoup de travail reste à faire en ce qui a trait à l’interopérabilité des logiciels et des protocoles, et à l’adoption de normes sur les métadonnées ainsi que sur la qualité et le partage des données.

Niveau d’écart

Modéré

VI. Compétences et formation

État idéal : *Les activités liées à la gérance des données sont répandues et appuyées par des scientifiques en matière de données et des professionnels de l’information spécialement formés. De plus, les chercheurs sont instruits en ce qui concerne les principes et l’importance de la gérance des données. Ils sont conscients de leurs propres rôles et responsabilités.*

État actuel

Aucune enquête détaillée n’a été réalisée quant aux niveaux de compétences et de formation. Cependant, la documentation indique que :

- de nombreux chercheurs ne connaissent pas les processus de gérance de données, y compris l’importance des métadonnées;
- peu de chercheurs ont reçu une formation précise sur le développement et la conservation des bases de données;
- de nombreux chercheurs sont réticents à assumer la responsabilité de la gérance des données au-delà de leurs intérêts immédiats.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Oui	Dans un nombre limité de cas – p. ex. ateliers du CNC/CODATA* pour les chercheurs	Faible niveau de compétences et peu de scientifiques spécialistes des données sont employés par les institutions pour appuyer les chercheurs	Dans un nombre limité de cas – p. ex. programme de formation de l’IDD*

*Voir le glossaire.

Écarts

Il manque de scientifiques et de professionnels de l'information formés, qui connaissent le catalogage des données, les normes et les processus en matière de métadonnées, la gestion de la conservation et l'évaluation de la valeur des données pour appuyer les chercheurs. Les gestionnaires de données ne sont pas encore considérés globalement comme étant essentiels à la recherche et demeurent vulnérables aux pressions budgétaires, surtout lorsque de tels « frais généraux de bibliothèque » exigent des augmentations budgétaires. De plus, le milieu de la recherche ne semble pas comprendre l'importance de la gestion des données, et il existe peu d'occasions pour les chercheurs de recevoir une formation sur les questions liées à la gestion des données.

Niveau d'écart

Important

VII. Systèmes de récompense et de reconnaissance

État idéal : *Les systèmes de récompense pour les chercheurs reconnaissent largement les contributions aux données de recherche et l'élaboration d'outils pour la gestion, l'utilisation et la conservation améliorées des données en tant qu'indicateurs importants de rendement.*

État actuel

- Les chercheurs sont évalués principalement selon leur registre de publications plutôt que selon les activités liées à la gestion des données.
- Le CRSH dispose d'une politique d'archivage des données, mais ne récompense pas les chercheurs qui se conforment à cette politique dans le cadre des demandes futures qu'ils peuvent présenter.
- Les organismes de financement et les universités disposent de politiques de partage des données, mais il n'est pas clair jusqu'à quel point la conformité est reconnue (ou la non-conformité est pénalisée).
- Quelques projets, tels que l'Année polaire internationale, exigent des plans de gestion des données de la part des chercheurs comme condition de financement.
- Dans certaines disciplines, il existe des publications qui présentent des numéros spéciaux axés sur les bases de données novatrices développées dans la collectivité. Certains documents publient également des études sur les bases de données (p. ex. *Journal of Bioinformatics*).

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Incitatifs intégrés liés à la réussite d'un projet de recherche	Dans les universités et les organismes de financement, certaines mesures	Dans un nombre limité de cas – certains projets de recherche, les organismes de financement et les	Information inconnue

	en place tendent à décourager les chercheurs qui préparent les données pour le partage.	documents exigent le dépôt des données dans des dépôts ou des services d'archivage. La non-conformité entraînerait probablement la perte de financement futur ou la non-publication.	
--	---	--	--

Écarts

Dans nombre de domaines, les cultures de recherche ne favorisent ni la conservation ni le partage des données. Par conséquent, les chercheurs ne comprennent pas les niveaux de risque liés à leurs pratiques courantes. Les systèmes de récompense pour les chercheurs ne reconnaissent pas les pratiques saines de gestion et de partage des données. Il y a peu de reconnaissance du leadership lié à la compilation de bases et d'ensembles de données de grande valeur et en libre accès, aux contributions d'envergure à ces bases et ensembles de données, et à l'élaboration d'outils qui améliorent la valeur des données (p. ex. l'élaboration de méthodes pour l'exploration en profondeur de données ou le regroupement de bases de données à l'échelle des disciplines). Les contributions importantes aux données de recherche et la gestion des données ne sont pas soulignées dans le cadre des examens de permanence ou de promotion des chercheurs au sein des établissements universitaires.

Niveau d'écart

Important

VIII. Recherche et développement

État idéal : *À l'appui des besoins en matière de gérance des données, le Canada a adopté une approche coordonnée des activités de recherche et de développement, dans le cadre de laquelle les priorités sont bien articulées et le financement adéquat.*

État actuel

- Au Canada, un certain nombre de projets non reliés examinent différentes questions liées à la gérance des données. Par exemple, le projet InterPARES (*International Research on Permanent Authentic Records in Electronic Systems* : <http://www.interpares.org/>), centralisé à l'Université de la Colombie-Britannique, contribue à la recherche internationale dans le domaine de la conservation et de l'authenticité des renseignements numériques, y compris les données.
- Les dépôts de données sont créés par chaque projet de recherche, mais ces solutions axées sur les projets ne peuvent habituellement pas être généralisées ni utilisées par d'autres.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage

Des dépôts de données axés sur les disciplines sont en voie de développement dans le cadre de projets de recherche particuliers afin de faciliter l'utilisation et l'analyse des données.	Information inconnue	Certains projets en voie de réalisation par le système InterPARES au Canada	Information inconnue
---	----------------------	---	----------------------

Écarts

De nombreux défis demeurent en matière de recherche et de développement dans le domaine de la gérance des données. Il existe un besoin de recherche continue sur les technologies, les modèles organisationnels, les normes et pratiques et l'interopérabilité, puisque le volume des données s'accroît et devient plus complexe. Les projets de recherche et de développement ne sont ni coordonnés ni orientés par des priorités nationales. Aucun effort coordonné n'est déployé pour participer à la recherche et au développement internationaux dans ce domaine.

Niveau d'écart

Modéré

IX. Accessibilité

État idéal : *L'accès aux données de la recherche financée par les deniers publics est étendu et les mécanismes en place pour le réglementer sont appropriés et tiennent compte des intérêts économiques, juridiques, de sécurité et d'éthique, s'il y a lieu.*

État actuel

- Une bonne partie des données de recherche produites aujourd'hui sont difficiles d'accès par les autres milieux de recherche canadiens et ne sont pas souvent idéalement structurées pour être aussi utiles ou ouvertes que possible, même à l'intérieur de la discipline pour laquelle elles ont été réalisées.
- Actuellement, d'importantes réserves de données sont inutilisées et ne sont pas disponibles en ligne.
- Les chercheurs sont réticents à partager les données parce qu'ils croient qu'elles font partie de leur propriété intellectuelle.
- Les chercheurs ne possèdent pas l'expertise nécessaire pour garantir que les données seront accessibles à autrui à l'avenir.
- Bien qu'il existe une tendance internationale croissante vers l'accès gratuit aux données conservées dans les dépôts (p. ex. GéoConnexions), beaucoup de dépôts de données facturent encore des frais

d'accès, par l'intermédiaire de mécanismes tels que les licences d'utilisation et le paiement à la carte, ou limitent l'accès aux membres de certaines collectivités (p. ex. Statistique Canada et Ressources naturelles Canada).

- Peu d'organismes de recherche disposent de politiques exigeant que les chercheurs fournissent l'accès aux données. La plupart des institutions de recherche exigent que les données soient conservées pendant cinq ans et partagées avec autrui sur demande. Lorsqu'elles existent, les politiques ne couvrent pas tous les types de données de recherche. Les politiques relatives à la protection des renseignements personnels et à l'éthique exigent que certains types de données soient anonymisés ou détruits.
- La priorité est donnée à l'utilisation à court terme à des fins d'analyse et de publication d'articles.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
<p>La priorité est axée sur l'analyse immédiate, et non sur la diffusion et la réutilisation par autrui.</p>	<p>Souvent, les chercheurs ne disposent ni du temps ni des compétences nécessaires pour la préparation des données. Les chercheurs sont réticents à partager les données. Les chercheurs ne savent pas à qui appartiennent réellement les données.</p>	<p>Il existe d'importants écarts dans la couverture et la capacité des dépôts de données.</p>	<p>En ce qui a trait à l'accessibilité, le paiement à la carte et les licences d'utilisation sont encore courants. La plupart des données sont conservées sur les disques durs des ordinateurs des chercheurs et ne sont pas accessibles à ceux qui souhaiteraient les utiliser. Beaucoup d'organismes qui sont propriétaires de données ne sont pas équipés pour fournir des services de diffusion, tels qu'une bonne documentation, des formats uniformisés, l'assurance du respect des règles de l'éthique et des exigences de confidentialité, la</p>

			communication des modalités et des conditions d'utilisation, la documentation du volume d'utilisation. Sans ces éléments, l'accessibilité en ligne est sans intérêt.
--	--	--	--

Écarts

Souvent, les chercheurs n'ont ni le temps ni les compétences nécessaires pour la préparation de la diffusion des données, sont réticents à partager les données et ne savent pas à qui appartiennent réellement les données. Il existe peu de mécanismes de récompense ou de reconnaissance pour le partage des données de recherche. Nombre d'organismes ne sont pas équipés pour fournir des services de diffusion, ce qui comprend la capacité de fournir une bonne documentation et des données dans des formats uniformisés, d'assurer le respect des règles de l'éthique et des exigences de confidentialité, de communiquer les modalités et conditions d'utilisation et de surveiller l'utilisation globale.

Niveau d'écart

Modéré

X. Conservation

État idéal : *Les données de recherche ayant une valeur durable sont conservées selon des pratiques actives de gestion axées sur les normes tout au long du cycle de vie des données. De plus, les données sont intégrées dans un environnement institutionnel durable qui appuie les dépôts de données numériques fiables.*

État actuel

- Peu d'organismes de recherche disposent de politiques de conservation des données de recherche.
- Lorsqu'elles existent, les politiques ne couvrent pas tous les types de données.
- La politique d'archivage des données actuelle du CRSH n'est pas exécutoire.
- Les bases de données du gouvernement du Canada, telles que celles qui sont maintenues par Statistique Canada, Pêches et Océans Canada, Ressources naturelles Canada et Environnement Canada, contiennent certaines collections de données importantes qui sont l'objet de bonnes mesures de conservation.
- Les organismes de financement de la recherche exigent rarement des plans de gestion des données.
- Un grand nombre de chercheurs n'ont le temps ni les compétences nécessaires pour la préparation de la conservation des données.

- La priorité est donnée à l'utilisation à court terme pour l'analyse des données et la publication d'articles.
- Il existe peu de mécanismes de récompense ou de reconnaissance pour la conservation des données de recherche.
- Lorsque des activités liées à la conservation existent, les pratiques les plus communément citées sont le contrôle de la qualité, l'entreposage et la sauvegarde des données, ainsi que les métadonnées descriptives.

Production des données	Diffusion des données	Gestion à long terme des données	Découverte et recyclage
Peu de chercheurs ont le temps ou les compétences nécessaires pour la préparation de la conservation des données.	Information inconnue	La couverture et la capacité des dépôts de données au Canada sont limitées. Le financement pour les dépôts de données est limité.	S.O.

Écarts

Les collections de données sont appuyées par des budgets relativement petits, souvent par l'intermédiaire de subventions de recherche qui financent un projet précis. Par conséquent, la priorité n'est pas accordée à leur conservation. La plupart des données recueillies grâce à la recherche ne sont pas déposées dans des dépôts. Peu ou aucun dépôt ou service d'archivage de données n'a la pleine capacité de conservation au sens du statut des dépôts de données numériques fiables. Il y a également un manque de sensibilisation dans les milieux de recherche relativement aux normes de conservation. Aucun organisme national ne fournit une orientation aux chercheurs quant aux normes. Le Canada continuera de perdre de précieuses données de recherche si aucun financement n'est accordé aux chercheurs pour la préparation des données et aux institutions pour la collecte des données et la conservation de ces dernières après l'achèvement d'un projet de recherche.

Niveau d'écart

Important

Conclusion et prochaines étapes

En raison du volume croissant de données de recherche, de nombreuses options existent pour l'utilisation et l'exploitation de ces ressources de façon à découvrir de nouvelles connaissances et fournir au Canada un avantage concurrentiel. Toutefois, sans la gestion appropriée de ces ressources numériques tout au long du cycle de vie des données, le Canada aura gaspillé cette occasion. Une évaluation de l'état de plusieurs indicateurs clés à l'échelle des étapes de ce modèle de cycle de vie a été effectuée afin de déterminer les

priorités et le travail nécessaires pour placer le Canada dans un environnement de données de recherche sain. Les résultats indiquent de sérieux écarts entre l'état actuel et l'état idéal dans le secteur de la recherche. Ces écarts sont résumés dans les tableaux 1 et 2.

Tableau 1 : Écarts à l'échelle du cycle de vie des données

Production des données

- La priorité est axée sur l'utilisation immédiate plutôt que sur les possibilités d'exploitation à long terme.
- Il existe un nombre limité de mécanismes de financement pour la préparation appropriée des données aux fins d'utilisation future.
- Peu d'institutions de recherche exigent des plans de gestion des données.
- Aucun organisme national ne peut fournir des conseils et aider à l'application des normes en matière de données.

Diffusion des données

- Il y a un manque de politiques pour régir le respect des normes appliquées afin d'assurer la diffusion des données.
- Les chercheurs sont réticents à partager les données, parce qu'ils ne disposent pas du temps et des compétences nécessaires.
- Certaines politiques exigent que certains types de données soient détruits après l'achèvement d'un projet de recherche.

Gestion à long terme des données

- Les dépôts de données ne disposent pas de la couverture et de la capacité nécessaires.
- Les activités liées à la conservation des renseignements dans les dépôts de données ne sont pas exhaustives.
- Au Canada, le financement des dépôts de données est limité.
- Il existe peu d'incitatifs invitant les chercheurs à déposer les données dans des archives.

Découverte et recyclage

- La plupart des données sont conservées sur les disques durs des ordinateurs des chercheurs et ne sont pas accessibles à autrui.
- Lorsque des données sont disponibles, les mécanismes d'accès par paiement à la carte et licence d'utilisation sont courants.
- Un grand nombre de chercheurs sont réticents à permettre l'accès à leurs données parce qu'ils croient qu'elles font partie de leur propriété intellectuelle.

Tableau 2 : Résumé de l'analyse des écarts

Indicateur	Niveau d'écart	
	→	
Politiques	Modéré	
Financement	Important	
Rôles et responsabilités	Important	
Dépôts de données [numériques fiables]	Important	
Normes	Modéré	
Compétences et formation	Important	
Systèmes de récompense et de reconnaissance	Important	
Recherche et développement	Modéré	
Accessibilité	Modéré	
Conservation	Important	

Un certain nombre de raisons sous-tendent ces écarts, notamment l'absence de politiques, d'infrastructure et de mécanismes de financement pour la gérance des données, ainsi qu'une culture de recherche qui ne reconnaît pas la valeur de la gestion des données. La gérance des données au Canada souffre de l'absence d'une vision stratégique nationale. L'approche fragmentée existante a entraîné de sérieux écarts à l'échelle du cycle de vie des données, ce qui est particulièrement évident dans les trois étapes finales. En raison de l'absence de services et de procédures pour la livraison des données à d'autres chercheurs, des quantités importantes de données sont rendues inaccessibles. Le nombre pathétiquement inadéquat de dépôts de données fiables au Canada contribue à l'écart déterminé à l'étape de la gestion à long terme des données. Par conséquent, la plupart des données de recherche créées au Canada sont grandement sous-utilisées et en grand danger d'être perdues.

Il existe des variations considérables dans les différentes disciplines, ce qui contribue à la complexité de ces questions. Les sciences sociales sont avancées dans certains domaines, en raison de la norme internationale de documentation des données de la Data Documentation Initiative (DDI), qui a été adoptée au Canada par l'Initiative de démocratisation des données. Dans le domaine des sciences, une distinction peut être faite entre les projets de « mégascience », qui recueillent d'énormes quantités de données qui sont systématiquement conservées, et les projets de « miniscience », qui tendent à générer des types très variés

de données qui, règle générale, ne sont ni facilement accessibles ni activement conservées. Entre-temps, la pratique d'archivage des données est relativement rare dans les domaines des arts et des sciences humaines.

Des risques importants sont associés à l'inaction. Les chercheurs canadiens ne disposeront pas des outils dont ils ont besoin pour demeurer à la fine pointe, et la capacité d'innovation du Canada finira par diminuer. Le Royaume-Uni et les États-Unis, entre autres, sont déjà beaucoup plus avancés en ce qui a trait à l'élaboration de politiques et d'infrastructures. Ils reconnaissent que les données constitueront un facteur déterminant dans les nouvelles découvertes. Par exemple, aux États-Unis, la National Science Foundation a récemment investi 100 millions de dollars dans l'élaboration d'un réseau de dépôts de données durable et interopérable pour appuyer la recherche scientifique.

Compte tenu de la portée de cet enjeu, il est clair que ces questions ne peuvent pas être résolues isolément. Elles doivent être résolues collectivement avec la participation de tous les secteurs et de toutes les disciplines des milieux de recherche. Les prochaines étapes pour le groupe de travail seront d'élaborer une stratégie concertée qui formulera le besoin d'une approche systématique de la gérance des données et qui déterminera également certains résultats atteignables à court et à moyen termes, afin que des progrès puissent être réalisés en ce qui a trait au rapprochement des écarts précisés ci-dessus.

Glossaire

CANARIE – CANARIE inc. : société qui facilite le développement et l'utilisation de son réseau, ainsi que des produits, des applications et des services évolués qui y sont exploités.

FCI – Fondation canadienne pour l'innovation : société indépendante créée par le gouvernement du Canada pour financer l'infrastructure de recherche.

IRSC – Instituts de recherche en santé du Canada : organisme de financement de la recherche en santé du gouvernement du Canada.

CNC/CODATA – Comité national canadien pour le CODATA (Comité pour les données scientifiques et technologiques) : la voix canadienne du CODATA, le comité pour les données scientifiques et technologiques. Le CODATA est un comité scientifique interdisciplinaire du Conseil international pour la science (CIS), qui a été établi il y a 40 ans.

IDD – Initiative de démocratisation des données : programme de Statistique Canada qui permet aux institutions participantes de payer des frais d'abonnement annuels pour la fourniture à leur faculté et aux étudiants d'un accès illimité aux nombreux fichiers de microdonnées, aux bases de données et aux fichiers géographiques pour utilisation publique de Statistique Canada. L'utilisation de ces fichiers est restreinte à la recherche universitaire et à l'enseignement.

API – Année polaire internationale : programme scientifique important axé sur l'Arctique et l'Antarctique de mars 2007 à mars 2009.

CRSNG – Conseil de recherches en sciences naturelles et en génie du Canada : organisme de financement de la recherche en sciences naturelles et en génie du gouvernement du Canada.

LPRPDE – *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE) : la loi canadienne sur la protection des renseignements personnels du secteur privé.

CRSH – Conseil de recherches en sciences humaines du Canada : un organisme fédéral qui favorise et appuie la recherche et la formation universitaires dans les domaines des sciences humaines et des sciences sociales.

Gérance des données de recherche au Canada : une analyse des écarts

Appendice

Tableau 1 : États idéaux

Politiques : Les organismes canadiens disposent de politiques cohérentes fondées sur des principes de gestion de données sains, qui régissent la gestion des données à l'échelle des disciplines tout au long du cycle de vie de la gérance des données et qui reflètent les initiatives législatives pertinentes.

Financement : Pris dans son ensemble, l'éventail des mécanismes de financement couvre les coûts tout au long du cycle de vie des données, ce qui assure un appui à long terme pour les nombreux intervenants de la recherche.

Attributions [Rôles et responsabilités] : Les attributions sont clairement définies et entièrement assumées. Chaque participant au cycle de vie des données assume un ensemble distinct de responsabilités et doit agir de concert avec d'autres participants dans la poursuite d'objectifs globaux et importants en matière de gérance des données pour l'ensemble de la collectivité.

Dépôts de données numériques fiables : Le Canada dispose d'un réseau complet de dépôts de données numériques fiables qui offrent un accès continu et fiable à toutes les données de recherche considérées comme ayant une valeur durable.

Normes : Des normes (objet, processus et instrumentation) sont adoptées de façon généralisée dans le cadre de chacune des étapes du cycle de vie des données et elles sont mises en œuvre indépendamment du domaine.

Compétences et formation : Les activités liées à la gérance des données sont répandues et appuyées par des scientifiques en matière de données et des professionnels de l'information spécialement formés. De plus, les chercheurs sont instruits en ce qui concerne les principes et l'importance de la gérance des données. Ils sont conscients de leurs propres attributions.

Systèmes de récompense et de reconnaissance : Les systèmes de récompense pour les chercheurs reconnaissent largement les contributions aux données de recherche et l'élaboration d'outils pour la gestion, l'utilisation et la conservation améliorées des données en tant qu'indicateurs importants de rendement.

Recherche et développement : À l'appui des besoins en matière de gérance des données, le Canada a adopté une approche coordonnée quant aux activités de recherche et de développement, dans le cadre de laquelle les priorités sont bien articulées et le financement adéquat.

Accessibilité : L'accès aux données de la recherche financée par les deniers publics est étendu et les mécanismes en place pour le réglementer sont appropriés et tiennent compte des intérêts économiques, juridiques, de sécurité et d'éthique, s'il y a lieu.

Conservation : Les données de recherche ayant une valeur durable sont conservées selon des

pratiques actives de gestion axées sur les normes en vigueur tout au long du cycle de vie des données. De plus, les données sont intégrées dans un environnement institutionnel durable qui appuie les dépôts de données numériques fiables.

Tableau 2 : Attributs des politiques en matière de données

Organisme	Portée	Accès aux données	Archives de données	Exigences
IRSC	Données sélectives – coordonnées bio-informatiques, atomiques et moléculaires	Dès leur publication	Base de données publique appropriée	Par ailleurs, tous les bénéficiaires de subventions doivent conserver les ensembles de données originaux pendant un minimum de cinq ans après la fin de la subvention; exigence qui s'applique à toutes les données, publiées ou non.
CRSH	Toutes les données de recherche rassemblées grâce au financement du CRSH	Dans les 2 ans suivant la fin du projet de recherche	Bibliothèque ou service de données de l'établissement, si l'archivage de données est possible. Sinon, liste des archives de données au Canada	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
Programme de subventions de réseaux stratégiques du CRSNG	Grands ensembles de données rassemblées grâce au financement du Programme de subventions de réseaux stratégiques du CRSNG	Dans des délais raisonnables	Aucune mention	Une entente établissant la responsabilité en matière de mise à jour et de conservation de grands ensembles de données doit être négociée dès le début des activités du réseau.
Énoncé de politique des trois Conseils sur l'accès aux renseignements confidentiels	Données assorties de renseignements personnels permettant d'identifier une personne	S. O.	Aucune mention	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 renseignements permettant d'identifier des personnes ou des groupes devraient être conservés dans des banques de données différentes, avec des identificateurs distincts.
Politique type	Toutes les	Données	Aucune	Les données doivent être

Organisme	Portée	Accès aux données	Archives de données	Exigences
d'une université (Université McGill)	données	préservées pendant 5 ans	mention	présentées de sorte qu'elles soient facilement vérifiées. Sous réserve d'exceptions établies en fonction de l'obligation de confidentialité et les lois relatives à la propriété intellectuelle et à l'accès à l'information, une fois les données publiées, celles-ci doivent être accessibles à toute partie soumettant une demande raisonnable pour les consulter.
Modèle type – Comité d'éthique de la recherche sur des sujets humains (Université de Victoria)	Recherche sur des sujets, restes et cadavres humains, des tissus et liquides biologiques, des embryons, des fœtus et tout autre matériel biologique, y compris l'ADN et l'ARN humains ou des fragments d'ADN ou d'ARN. Nécessite l'autorisation ou la renonciation du Comité avant d'entreprendre les travaux de recherche.	S. O.	Aucune mention	<p>Les chercheurs doivent fournir des plans de conservation ou d'élimination des données des participants qui sont appropriés pour le domaine de recherche en cause et qui respectent la volonté des participants. Par exemple, pour ce qui est de l'histoire orale, la meilleure pratique pourrait être d'archiver l'information recueillie (avec le consentement des participants) pour les générations futures. Quant à la recherche où la diffusion d'information pourrait porter préjudice aux participants, il serait peut-être préférable de détruire les données recueillies le plus rapidement possible.</p> <p>Détaillez vos plans pour ce qui est de la conservation et la protection des données des participants ou de leur destruction en tenant compte des meilleures pratiques dans votre domaine de recherche et de la volonté des participants. Certains organismes de financement, organisations professionnelles et éditeurs ont établi des exigences minimales en matière de conservation des données (p. ex., 5 ans), après quoi elles doivent être détruites. Vous devez divulguer leurs plans concernant l'élimination des données qui font état d'un échéancier et des méthodes qui seront utilisées pour détruire les données (p. ex., déchiquetage et suppression de fichiers)</p>

Organisme	Portée	Accès aux données	Archives de données	Exigences
				électroniques).
Politique type – ministère fédéral (Pêches et Océans Canada)	L'ensemble des données scientifiques du MPO	Dans les 2 ans suivant leur acquisition	La gestion des archives – Service des données sur le milieu marin (SDMM), secteur Sciences, assurera au besoin la coordination entre les centres régionaux, zonaux et nationaux afin de voir à ce que toutes les données soient gérées de façon appropriée. Si aucun centre de données n'existe dans une région, les gestionnaires des sciences et des océans devront désigner et supporter des postes dotés pour une période indéterminée à l'aide des ressources des services votés qui comprendront les responsabilités associées à la gestion des données.	<p>Pour assurer la gestion et l'archivage appropriés des données, toutes les données scientifiques recueillies par le ministère doivent être transférées dans des archives « gérées » immédiatement après leur traitement.</p> <p>Pour offrir un maximum d'avantages au ministère et à la communauté d'utilisateurs, les données scientifiques doivent être accessibles au moment opportun, sans restriction et en toute liberté, conformément aux obligations ministérielles, nationales et internationales à l'égard de son fonds de données.</p>
Génome Canada	Non précisée	Au plus tard au moment de l'acceptation pour publication des principaux	Génome Canada donne des exemples de bases de données dans	Le partage des données doit se faire en temps opportun. Génome Canada s'attend à ce que les données soient diffusées et partagées au plus tard au moment

Organisme	Portée	Accès aux données	Archives de données	Exigences
		résultats découlant d'un ensemble de données issues d'un projet	lesquelles différents types de données ou de ressources uniques produites par les projets financés par Génome Canada peuvent être déposés.	de l'acceptation pour publication des principaux résultats découlant d'un ensemble de données issues d'un projet. Dans le cas des grands ensembles de données qui sont recueillies sur plusieurs périodes ou en plusieurs phases, il est raisonnable de s'attendre à ce que les données soient publiées par phase, lorsqu'elles deviennent disponibles ou que les principaux résultats d'une phase de la recherche sont publiés. Toutes les données doivent toutefois être publiées sans restriction à la conclusion d'un projet.
Modèle type – politique relative à un projet de recherche (Année polaire internationale – API)	Les données de l'API sont les données générées durant le calendrier de l'API (du 1 ^{er} mars 2007 au 1 ^{er} mars 2009) lors de projets particuliers reconnus comme étant des projets de l'API par le comité mixte de l'ICSU/WMO.	Dans les plus brefs délais	Le service d'information et de données de l'API devrait collaborer avec les centres opérationnels, les centres de données et les autres organismes pertinents afin d'assurer la préservation des données qui sont connexes à l'API et qui ne sont pas explicitement issues de ses projets.	<p>Selon le comité mixte de l'API, les données de l'API, y compris les données opérationnelles, doivent non seulement être fournies en temps réel, mais être aussi accessibles dans leur intégralité, gratuitement et en libre accès, dans les plus brefs délais.</p> <p>Puisque l'attestation de la véritable valeur des données scientifiques est généralement établie longtemps après leur collecte, et pour assurer la pérennité patrimoniale de l'API, il est essentiel de garantir la conservation à long terme et l'accès continu aux données de l'API. Toute donnée de l'API doit être archivée de la manière la plus simple et la plus pratique possible, et accompagnée d'une description complète des métadonnées.</p>
Loi sur la protection des renseignements personnels	Renseignements personnels	S. O.	S. O.	Une institution fédérale procède au retrait des renseignements personnels qui relèvent d'elle, conformément aux règlements, aux instructions ou aux directives applicables du ministre désigné.
Loi sur la protection des renseignements	Renseignement personnel – tout renseignement	Les données doivent être éliminées dès	S. O.	Les renseignements personnels seront utilisés ou divulgués uniquement en vue d'atteindre les

Organisme	Portée	Accès aux données	Archives de données	Exigences
personnels et les documents électroniques	concernant une personne identifiable, à l'exclusion du nom et du titre d'un employé d'une organisation et de l'adresse et du numéro de téléphone de son lieu de travail	qu'elles ne sont plus nécessaires à l'atteinte des objectifs précisés		<p>objectifs pour lesquels ils ont été recueillis, sauf avec le consentement de la personne ou selon les exigences de la loi. Les renseignements personnels ne seront conservés que pour la période nécessaire à l'atteinte de ces objectifs.</p> <p>Lorsque les renseignements personnels ne seront plus nécessaires à l'atteinte des objectifs en vue desquels ils ont été recueillis, ils seront détruits, effacés ou rendus anonymes. Les organismes doivent élaborer des lignes directrices et appliquer des procédures régissant la destruction des renseignements personnels.</p>
Législation portant sur le dépôt légal	Ce ne sont pas tous les documents en ligne qui sont assujettis à la portée de la législation sur le dépôt légal. BAC se concentre sur les documents en ligne qui sont considérés comme sous une forme de « publication ». Les publications en ligne ont habituellement un titre distinct, un auteur particulier ou une autorité auteure ainsi qu'une date particulière et sont destinées à être lues par le public. Les catégories de publications en	Dès leur publication	Bibliothèque et Archives Canada	Bibliothèque et Archives Canada (BAC) a pour mandat de préserver le patrimoine documentaire du pays pour les générations actuelles et futures. La Loi sur la Bibliothèque et les Archives du Canada a été adoptée en avril 2004 et, en vertu du Règlement sur le dépôt légal, BAC peut bâtir et conserver une collection complète du patrimoine documentaire du Canada.

Organisme	Portée	Accès aux données	Archives de données	Exigences
	<p>ligne qui devraient être déposées sont notamment les livres, magazines, rapports annuels, mémoires de recherche et revues savantes. Les catégories de publication en ligne qui ne devraient pas être déposées sont, entre autres, les formulaires, correspondances électroniques, résumés analytiques, communiqués de presse, portails, annonces, horaires, calendriers et bases de données.</p>			

Tableau 3 : Sources et bénéficiaires de financement pour les activités de gérance de données tout au long du cycle de vie

Nota : Cette liste n'est pas exhaustive.

Source de financement	Étape de la production	Étape de la diffusion	Étape de la gestion	Étape de la découverte et du recyclage
Trois conseils Programme de subventions à la découverte – IRSC, CRSNG, CRSH	Équipes de recherche			
CRSH – « Les dépenses relatives à la préparation des données de recherche à archiver sont admissibles aux programmes de subventions de recherche du CRSH. »		Équipes de recherche	Équipes de recherche	
FCI – Selon les lignes directrices relatives au financement, la FCI accepte de financer l'acquisition de bases de données ou le processus de conception et de développement de bases de données jusqu'au point où ces bases de données peuvent être exploitées par les chercheurs visés par le projet.	Équipes de recherche			
FCI – Fonds d'exploitation des infrastructures : « Coûts du personnel technique et d'exploitation directement liés à l'exploitation et à l'entretien de l'infrastructure (p. ex., coût d'un technicien chargé d'entretenir et de faire fonctionner			Équipes de recherche	

Source de financement	Étape de la production	Étape de la diffusion	Étape de la gestion	Étape de la découverte et du recyclage
l'infrastructure). »				
CRSH, IRSC, FCI Financement des centres de données de recherche		Établissements		Établissements
Programme des coûts indirects	Établissements (« Les ressources d'information comprennent les bases de données, les télécommunications, les technologies de l'information et les outils de recherche. »)		Établissements – Coûts d'exploitation de bibliothèque ou de centre de données	
CANARIE (Plateforme réseautée) – Les outils pour l'acquisition, la conservation, la manipulation, la mutualisation et l'analyse des données utilisés par la communauté repartie sont d'un intérêt direct dans le cadre de ce programme.	Équipes de recherche	Équipes de recherche	Équipes de recherche	Équipes de recherche
Les organismes provinciaux de financement (p. ex., Fonds pour la recherche en Ontario)	Établissements – FRO – Le Programme d'infrastructure de recherche « finance les coûts d'immobilisations, de l'acquisition, de la conception, de la modernisation ou de la location-bail des coûts admissibles totaux. »			Établissements – FRO – Programme d'excellence en recherche – les coûts indirects, c'est-à-dire, les frais généraux du projet de recherche

Source de financement	Étape de la production	Étape de la diffusion	Étape de la gestion	Étape de la découverte et du recyclage
Conseil des bibliothèques universitaires de l'Ontario et AchatsOntario				Consortiums de bibliothèques (OCUL) – Ontario Data Documentation, Extraction Service and Infrastructure Initiative (ODESI)
Gouvernement provincial et fédéral	Ministères : Agriculture et Agroalimentaire Canada; Agence spatiale canadienne/Institut Herzberg; Environnement Canada; Pêches et Océans Canada; GéoConnexions; Santé Canada; Conseil national de recherches Canada; Ressources naturelles Canada; Statistique Canada; divers ministères provinciaux	Ministères : Agriculture et Agroalimentaire Canada; Agence spatiale canadienne/Institut Herzberg; Environnement Canada; Pêches et Océans Canada; GéoConnexions; Santé Canada; Conseil national de recherches Canada; Ressources naturelles Canada; Statistique Canada; divers ministères provinciaux	Ministères : Agriculture et Agroalimentaire Canada; Agence spatiale canadienne/Institut Herzberg; Environnement Canada; Pêches et Océans Canada; GéoConnexions; Santé Canada; Conseil national de recherches Canada; Ressources naturelles Canada; Statistique Canada; divers ministères provinciaux	Ministères : Agriculture et Agroalimentaire Canada; Agence spatiale canadienne/Institut Herzberg; Environnement Canada; Pêches et Océans Canada; GéoConnexions; Santé Canada; Conseil national de recherches Canada; Ressources naturelles Canada; Statistique Canada; divers ministères provinciaux
Bibliothèque et Archives Canada			Collection d'un nombre restreint d'ensembles de données de recherche qui sont conformes à la définition de « publications ». Ces ensembles ne sont, toutefois, que de petits sous-ensembles de données de	

Source de financement	Étape de la production	Étape de la diffusion	Étape de la gestion	Étape de la découverte et du recyclage
			recherche devant être conservés au Canada.	
Universités	Laboratoires		Centres de données, dépôts institutionnels	Centres de données, Initiative de démocratisation des données
Organismes caritatifs de recherche	Laboratoires		Dépôts de données	

Tableau 4 : Responsabilités actuelles sur le plan de la gérance des données au Canada

	Production	Diffusion	Gestion à long terme	Découverte et recyclage
Chercheurs	Respecter les normes des règles de l'art	Travailler les données afin qu'elles puissent être utilisées par d'autres	Gérer les données au cours du cycle de vie du projet. Conserver les données pendant un certain temps	
Établissements			Gérer les données au cours du cycle de vie du projet	
Organismes de financement	Fournir des fonds pour la création de données (IRSC, CRSNG, CRSH) Élaborer des outils de recherche (CANARIE)		Fournir des fonds pour la gestion des données (FCI)	Fournir des fonds pour la conservation (coûts indirects)
Bibliothèques de recherche et bibliothèques de données			Préserver quelques ensembles de données uniques.	Fournir des outils pour la réutilisation des données
Ministères fédéraux	Entreprendre des activités de recherche	Fournir des outils pour la réutilisation des données créées à l'interne.	Gérer les données au cours du cycle de vie du projet. Gérer à long terme des ensembles de données particuliers.	
Bibliothèque et Archives Canada			Conserver un nombre restreint d'ensembles de données de recherche qui sont conformes à la définition de « publications ». Ces ensembles ne sont, toutefois, que de petits sous-ensembles de données de recherche devant être conservés au Canada.	

Tableau 5 : Liste de dépôts selon le modèle organisationnel

Nota : Cette liste n'est pas exhaustive.

<p>Archives de données provinciales ou fédérales – Ce type de dépôt recueille des données qui sont perçues comme revêtant une pertinence provinciale ou nationale. Dans certains cas, le dépôt a également le mandat de faciliter la science et la recherche au Canada. Les dépôts sont spécialisés et ont généralement une portée passablement restreinte pour ce qui est des types de données recueillies.</p>	<p>Agence spatiale canadienne/Institut Herzberg Agriculture et Agroalimentaire Canada Conseil national de recherches Canada Divers ministères provinciaux Environnement Canada GéoConnexions Pêches et Océans Canada Ressources naturelles Canada Santé Canada Statistique Canada</p>
<p>Organisme virtuel – Faciliter le partage des données pour la science électronique, qui, selon Tony Hey, est une « méthodologie de recherche guidée par les données » [<i>data-driven research methodology</i>]. (Nota – certaines archives de données fédérales visent également à faciliter la science électronique)</p>	<p>Venus Data Management and Archiving System Le projet Neptune Le projet Atlas L'Observatoire de neutrinos de Sudbury Génome Canada La base de données sur l'imagerie cérébrale de l'Institut neurologique de Montréal Eucalyptus TRIUMF</p>
<p>Centres de recherche universitaires – La gestion de ces dépôts est assurée par les départements universitaires (certains chevauchements avec des organismes virtuels)</p>	<p>Androgen Receptor Mutations Database Archives de données de l'Institute for Social Research Archives de données de l'International Infectious Disease Base de données d'affinité de fixation de l'hémoglobine Base de données de densité des molécules d'électrons organiques halogénées Base de données de densité du groupe d'électrons fonctionnel des composés carbonylés carcinogènes Base de données sur la toxicité aquatique des hydrocarbures aromatiques polycycliques Base de données sur la toxicité des métaux [Cadmium and Zinc Uptake by Grain Varieties Databank] Calcium Sensing Receptor Locus Mutation Database Cambridge Structural Database Canine Inherited Disorders Database Centre canadien de rayonnement synchrotron Centre des pêches de l'Atlantique nord-ouest CRYSYS (Système cryosphérique pour la surveillance des changements planétaires au Canada) Data for Evaluating Learning in Valid Experiments David Dunlop Observatory Database of Galactic Classical Cepheids Formulation analytique interactive en thermodynamique Fungal Mitochondrial Genome Genomics and proteomics Advanced Applications Project GOBASE – Organelle Genome Database HumGen Journal of Applied Econometrics Data Archive McMaster Cepheid Photometry and Radial Velocity Data Archive Organelle Genome Database (GOBASE)</p>

	<p>Organelle Genome Megasequencing Program (OGMP) Phenylalanine Hydroxylase Locus Knowledgebase Programme de l'ICRA en biologie évolutive (CIAR-PEB) PROMISE Software Engineering Repository Protist EST Program Site Web HEXdb - GM2 Gangliosidase Database Super Dual Auroral Radar Network (SuperDARN). The McGill Radar and Weather Data Archive University of Waterloo Weather Station Data Archive Viral Bioinformatics Resource Center Wilson Disease Mutation Database</p>
<p>Autres centres de recherche – Dépôts gérés par d'autres types de centres de recherche</p>	<p>Arabic Genetic Disease – Toronto Sick Kids Hospital Autism Chromosome Rearrangement Database Biological General Repository for Interaction Datasets (BioGRID) – Mount Sinai Hospital Réseau canadien des codes à barres ADN Chromosome 7 Annotation Project Cystic Fibrosis Mutation Database – Toronto Sick Kids Hospital Expression Profiles for C. elegans GFP:promoter Fusions Genome Sequence Centre Human Genome Segmental Duplication Database Non-Human Segmental Duplication Database Database of Genomic Variants The Lafora Progressive Myoclonus Epilepsy Mutation and Polymorphism Database National Database of FASD and Substance Use During Pregnancy Resources Pseudomonas Genome Database V2</p>
<p>Bibliothèques de données – Bien que le mandat principal soit de fournir un accès aux collections externes, quelques bibliothèques font la collecte d'ensembles de données de certaines disciplines, y compris, notamment, les sciences sociales.</p>	<p>University of Alberta Data Library University of British Columbia Data Library Carleton University Social Science Data Archive University of Guelph Data Resource Centre Queen's University Social Science Data Centre Simon Fraser University Research Data Library University of Toronto Data Library University of Waterloo Leisure Study Bank University of Waterloo Data Resource Centre University of Western Ontario Data Resource Library York University Institute for Social Research</p>
<p>Dépôts institutionnels – sous la gestion des bibliothèques de recherche, ces dépôts sont très nouveaux. Ils sont axés essentiellement sur la collecte de publications. Toutefois, peu de dépôts font la collecte de petites collections de données (p. ex., des images et très rarement des données numériques)</p>	<p>Centre de recherches pour le développement international Institut canadien de l'information scientifique et technique Université Carleton Université Dalhousie Université de Calgary Université de Guelph Université de l'Alberta Université de l'Île-du-Prince-Édouard Université de la Colombie-Britannique Université de la Saskatchewan Université de Lethbridge Université de Montréal Université de Regina Université de Toronto</p>

	Université de Victoria Université de Waterloo Université de Windsor Université de Winnipeg Université du Manitoba Université du Nouveau-Brunswick Université du Québec à Montréal Université Laval Université McGill Université McMaster Université Queen's Université Simon Fraser Université York
--	---

Tableau 6 : Barrières à l'accès aux données de recherche et à leur conservation au Canada

	Barrières
1. Politiques	<ul style="list-style-type: none"> - Ce ne sont pas tous les organismes qui disposent de politiques. - Les politiques ne couvrent pas toutes les données de recherche. - Certaines politiques sont contradictoires (« respect de la vie privée et éthique » et « accès et conservation »). - Les politiques ne font pas l'objet de surveillance ou de suivi. - Peu de politiques abordent la question de conservation à long terme.
2. Financement	<ul style="list-style-type: none"> - Petit nombre de mécanismes de financement disponibles pour la diffusion des données et leur conservation à long terme. - Consultation à la carte et licence d'accès encore très répandues.
3. Rôles et Responsabilités	<ul style="list-style-type: none"> - Les attributions visant à assurer l'accès ne sont pas clairement définies ou acceptées. - La responsabilité inhérente à la conservation des données est assumée uniquement par les chercheurs. - L'engagement des établissements envers la conservation à long terme des données demeure timide. - Les organismes de recherche n'exigent que rarement des plans de gestion des données.
4. Dépôts de données	<ul style="list-style-type: none"> - Absence de dépôts à l'appui de l'accès aux données et de leur conservation. - Les dépôts n'ont aucune capacité en matière de conservation.
5. Normes	<ul style="list-style-type: none"> - Faible conscientisation à l'égard des normes d'accès et de conservation. - Aucun organisme national pour orienter les chercheurs en ce qui a trait aux normes.
6. Compétences et formation	<ul style="list-style-type: none"> - Manque de scientifiques en matière de données et de professionnels de l'information pour assurer la conservation à long terme des données. - Les chercheurs croient vraisemblablement qu'ils n'ont ni le temps ni les compétences pour préparer les données.
7. Récompense et reconnaissance	<ul style="list-style-type: none"> - Les chercheurs sont réticents à partager leurs données. - Les chercheurs ne savent pas vraiment à qui appartiennent les données. - La priorité est l'utilisation à court terme des données à des fins d'analyse et de publication d'articles. - Petit nombre de mécanismes de récompense et de reconnaissance pour le partage des données de recherche.
8. Recherche et développement	<ul style="list-style-type: none"> - Les activités de R-D au Canada ne sont pas coordonnées et orientées par des priorités nationales.