CNADRS

Consultation nationale sur l'accès aux données de la recherche scientifique

Rapport final

31 janvier 2005

David F. Strong
et Peter B. Leach

CNADRS

Consultation nationale sur l'accès aux données de la recherche scientifique

Gouvernement du Canada
Government of Canada

Canada

Pour de plus amples renseignements ou pour obtenir une copie de ce rapport, veuillez vous adresser au :

Bureau d'aide à la clientèle de l'ICIST

Institut canadien de l'information scientifique et technique
Conseil national de recherches Canada
M-55, 1200, chemin Montréal
Ottawa (Ontario), Canada K1A 0R6

Tél. : 1-800-668-1222 (Canada et États-Unis) ou (613) 998-8544
Téléc. : (613) 993-7619
Courriel : info.icist@nrc-cnrc.gc.ca
URL : http://www.icist.cnrc.gc.ca

Consultation nationale sur l'accès aux données de la recherche scientifique

Rapport final
31 janvier 2005

David F. Strong
et Peter B. Leach

© Sa Majesté la Reine du Chef du Canada, 2005

Catalogage avant publication de Bibliothèque et Archives Canada

Canada. Groupe de travail de la Consultation nationale sur l'accès aux données de la recherche scientifique

Consultation nationale sur l'accès aux données de recherche scientifique : rapport final, janvier 31, 2005

Publ. aussi en anglais sous le titre : Report of the Consultation on Access to Scientific Research Data.
Publ. par le Groupe de travail de la Consultation nationale sur l'accès aux données de la recherche scientifique.
Disponible aussi sur l'Internet.
ISBN 0–662–79595–4
No de cat. NR16-84/2005F

1. Recherche – Canada – Documentation, Services de.
2. Sciences – Canada – Documentation, Services de.
3. Politique scientifique et technique – Canada.
4. Information scientifique – Politique gouvernementale – Canada.
6. Systèmes d'information – Recherche.
I. Institut canadien de l'information scientifique et technique.
II. Titre.

Q180.C3T37 2005 001.4'0971 C2005-980109-3

Papier – français : NR16-84/2005F, 0-662-79595-4
PDF – français : NR16-84/2005F-PDF, 0-662-79596-2

Table des matières

Avant-props v
Remerciements vi
1 Sommaire 1
2 Recommandations 5
Responsabilités du Groupe Données 5
Responsabilités de Données Canada 6
Responsabilités des organismes de financement 8
Responsabilités des universités et des chercheurs 9
3 Objectifs 11
4 Contexte 13
Introduction 13
Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics 15
Consultation sur les archives nationales de données du CRSH 16
5 Vision 19
6 Occasions et répercussions 21
Énoncé des répercussions sur dix ans 21
Nouvelle science 22
Meilleures connaissances scientifiques 23
Leadership en innovation 24
Politique et stratégie supérieures 25
Recherche plus efficace 25
Amélioration de l'éducation 26
7 Défis de l'accès libre 29
Priorité des besoin 29
Champions du changement 30
Culture 30
Formation 31
Normes et processus 31
Expertise en matière d'archivage 31
Responsabilités, systèmes et outils 32
Autres défis et occasions 33
8 Recommandations émanant de la consultation 35
Responsabilités du Groupe Données 35
Responsabilités de Données Canada 38
Responsabilités des organismes de financement 43
Responsabilités des universités et des chercheurs 46
9 Carte routière vers l'accès 49

10 Annexes A1
10.1 Cadre pour une vision pour 2010 – Amélioration de l'accès aux données de la recherche scientifique (novembre 2004) A1
10.2 OECD Final Report March 2003 (without appendices) A31
10.3 Rapport final de la Consultation sur les archives nationales de données - Juin 2002 (sans les annexes) A53
10.4 Déclaration sur l'accès aux données de la recherche financée par des fonds publics A83
11 Addenda A87
11.1 Sigles et acronymes A87
11.2 Ordre du jour de la consultation (sommaire fourni aux participants) A88
11.3 Participants à la consultation A90
11.4 Membres du Groupe de travail A92
11.5 Membres du groupe de gestion de projet. A93

Avant-propos

Au nom du Groupe de travail de la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS) et des personnes qui y ont participé, j'ai l'honneur de présenter ce rapport final. Je tiens également à remercier le Conseil national de recherches Canada (CNRC), la Fondation canadienne pour l'innovation (FCI), les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie (CRSNG) pour l'idée qui a donné naissance à la Consultation, pour l'engagement à l'égard de sa réussite et pour le soutien financier qui en a permis l'exécution.

La CNADRS constate l'urgente nécessité d'agir afin de propulser le Canada dans un nouveau contexte évolutif axé sur les données pour la recherche nationale. Nos recommandations proactives permettront au Canada de demeurer à l'avant-garde de la recherche mondiale et les progrès économiques, industriels, sociaux, environnementaux, écologiques et technologiques qui en résulteront permettront d'améliorer notre compétitivité mondiale, de rehausser la qualité de vie de tous les Canadiens et d'aider le pays à s'attaquer aux menaces de dégradation environnementale et de dommage écologique.

Je désire encourager l'étude immédiate et urgente de notre rapport et recommande la mise en œuvre dans les meilleurs délais, en tant que priorité nationale, de l'approche par étapes proposée qui aboutira à l'application prompte et efficace d'un plan national permettant de donner libre accès aux données de la recherche scientifique financée par l'État.

David Strong, Ph.D., DSc, LLD, FRSC
Président
Groupe de travail de la CNADRS

(Nota : Ce rapport a été rédigé en grande partie et avec talent par Peter B. Leach, président, Leach Technologies Ltd.)

Remerciements

En tant que président du Groupe de travail de la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS), je tiens à exprimer ma gratitude à tous ceux qui y ont participé, depuis la conception jusqu'à l'achèvement. Le soutien, l'aide et l'engagement sans faille du Groupe de travail de la CNADRS et des membres du Groupe de gestion de projet, de nos organisations partenaires et de leur personnel, ainsi que des participants au forum en ont garanti la réussite. Le Canada a enregistré des progrès en vue de garantir un accès libre aux données de la recherche scientifique financée par l'État et leur conservation à long terme et dans cette démarche, la contribution des organisations et des acteurs suivants mérite d'être reconnue. Donc, nous tenons à remercier :

- ❑ Tous les participants du forum et les organisations qu'ils représentent (voir la section 11.3).
- ❑ Tous les membres du Groupe de travail de la CNADRS (voir la section 11.4).
- ❑ Tous les membres du Groupe de gestion de projet de la CNADRS (voir la section 11.5).

Des remerciements spéciaux vont à l'Institut canadien de l'information scientifique et technique (ICIST) du Conseil national de recherches Canada (CNRC), hôte et organisateur de la CNADRS. Parmi le personnel de l'ICIST impliqué dans la consultation, nous aimerions mentionner Bernard Dumouchel, directeur général, et son adjointe administrative, Marie Charette; le groupe de Recherche de l'ICIST, notamment l'ancien chef de groupe Gordon Wood et son adjointe administrative, Carolyn Ahern, ainsi que Glen Newton, chef, Unité de la recherche; ainsi que l'équipe des Communications de l'ICIST, notamment Catherine Betz, Linda Campeau, George Guillemette, Stéphane Levesque et Alexandra Talbot. Jac van Beek et Stephanie Delorme des Services intégrés du CNRC ont également apporté une contribution importante à cette initiative.

Au nom de tous les participants, je tiens à remercier personnellement les conférenciers principaux : Arthur Carty, conseiller national des sciences auprès du premier ministre; Patricia Kosseim, ancienne directrice, Bureau de l'éthique, Instituts canadiens de recherche en santé; Claire Morris, présidente, Association des universités et collèges du Canada; Eliot Phillipson, président-directeur général, Fondation canadienne pour l'innovation; Steve Shugar, directeur, Politique et relations internationales, Conseil de recherches en sciences naturelles et en génie (CRSNG); Marie Tobin, ancienne directrice générale, Direction générale de la politique d'innovation, Industrie Canada; et Ian Wilson, bibliothécaire et archiviste du Canada. Vos connaissances et votre vision ont permis un débat constructif lors du forum de deux jours.

Enfin, je remercie tous les partenaires de cette initiative :

IRSC CIHR
Instituts de recherche en santé du Canada Canadian Institutes of Health Research

CRSNG NSERC

Fondation canadienne pour l'innovation
Canada Foundation for Innovation

CNRC-NRC

La présentation de ce rapport marque la fin de la CNADRS et le début du processus devant mener à la réalisation de ses recommandations. À tous ceux qui nous ont montré la voie de l'avenir, que ce soit sous les feux de la rampe ou en coulisse, je désire exprimer ma plus sincère reconnaissance.

Veuillez agréer mes salutations respectueuses,

David Strong, Ph.D., DSc, LLD, FRSC
Président
Groupe de travail de la CNADRS

1 Sommaire

Il y a près de 50 ans, le grand écrivain d'anticipation anglais, H. G. Wells, a prédit à peu de choses près l'avènement de l'Internet et du World Wide Web ainsi que les répercussions qu'ils pourraient avoir sur les connaissances et la recherche.

> *(Rares sont les personnes, étrangères à l'univers des bibliothécaires experts et des conservateurs de musée et ainsi de suite, qui savent combien il est facile de rendre maniables des faits soigneusement ordonnés, pour nombreux qu'ils soient, et combien il est facile de retrouver, rapidement et intégralement, les visions les plus rares et les questions les plus abstruses lorsqu'elles ont été mises à leur place dans un ensemble bien ordonné de référence et de reproduction.)*
>
> — H. G. Wells
> « World Brain: The Idea of a Permanent World Encyclopaedia »
> Contribution to the new Encyclopédie Française
> Août 1937

Ce jour est arrivé, et le Canada doit profiter de l'occasion.

À la mi-juin, un groupe d'experts nommés par le Conseil national de recherches Canada (CNRC) s'est réuni à Ottawa pour planifier un forum national devant servir de point de ralliement à la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS). Le forum a réuni plus de 70 chefs de file de la recherche, de la gestion des données, de l'administration, de la propriété intellectuelle et d'autres domaines importants en provenance de partout au Canada.

Ce rapport est un examen exhaustif des enjeux, des occasions et difficultés identifiés lors du forum, complété par une sélection de documents d'appui présentés en annexe.

1.1 Le nouveau monde

Un réseau riche et complexe de bases de données scientifiques change la façon dont la recherche est effectuée, accélère les possibilités de découverte et crée de nouveaux concepts. L'augmentation de l'accès à ce réseau accélèrera cette évolution et créera un nouveau monde de la recherche **et même**, **un tout nouveau monde**. Lorsque ces bases de données sont combinées à l'intérieur des disciplines et des pays et entre ceux-ci, des bonds prodigieux de la connaissance peuvent se produire et transformer notre compréhension de la vie, du monde et de l'univers.

Ainsi, dans l'analyse de la génétique humaine, la technologie qui permet de saisir des quantités énormes de données et de les analyser afin d'y découvrir de nouvelles informations permet déjà de révéler la constitution génétique de la vie et de mieux comprendre de nombreuses maladies et syndromes. Nous serons bientôt en mesure d'analyser des éléments d'une grande complexité, tels que la prédisposition des populations animales et végétales à la maladie en fonction de conditions génétiques, sociales, environnementales et démographiques, de façon à ce que tous ces facteurs puissent être intégrés à de nouvelles stratégies de prévention. La capacité d'avoir accès à des données compilées dans des domaines différents et de pouvoir les intégrer nous ouvre des secteurs entièrement nouveaux de la connaissance d'une façon qu'il aurait été impossible de concevoir dans le passé.

1.2 Le Canada

Pour que le Canada devienne un chef de file de l'économie du savoir, il doit être un chef de file du nouveau monde de la recherche. Pour que le Canada soit en tête de cette transformation de la recherche, il est essentiel qu'il donne promptement suite aux recommandations de ce rapport. Pour que le Canada puisse se développer sur le plan économique et social, il doit apporter des changements profonds au secteur scientifique, notamment :

- la culture et le comportement en matière de recherche;
- la gestion, les politiques et les stratégies des instituts de recherche;
- les cadres juridique et stratégique;
- le financement et la budgétisation de la recherche;
- les technologies de l'information et l'infrastructure informatique.

Quelques-uns de nos concurrents, membres de l'OCDE, réagissent beaucoup plus rapidement à ces défis et constituent une menace, voire un risque réel pour le bien-être économique et social du Canada.

Le Canada est certes impliqué dans de nombreuses bases de données et dans des initiatives mondiales de recherche, mais il s'agit de cas individuels qui ne sont pas reliés par une quelconque stratégie ou par des normes nationales. Par conséquent, une grande partie des données sur lesquelles reposent actuellement nos connaissances sont difficiles d'accès pour d'autres groupes de chercheurs canadiens et ne sont souvent pas structurées de façon idéale pour être aussi utiles ou aussi accessibles que possible, même dans les disciplines pour lesquelles elles ont été élaborées. Le fer de lance de l'activité nationale et de la présence internationale du Canada en matière d'accès aux données scientifiques est formé par le leadership combiné du Comité national canadien de CODATA (CNC/CODATA), de l'Institut canadien de l'information scientifique et technique (ICIST) et de l'Association des bibliothèques de recherche du Canada (ABRC).

Les institutions membres de l'ABRC s'emploient déjà à sauvegarder une partie du patrimoine scientifique du pays sous forme numérique et fournissent une grande partie des connaissances qui sont utilisées dans les processus de saisie, d'accès et de conservation de bon nombre des projets de bases de données scientifiques nationaux et internationaux. Les membres de l'ABRC sont vingt-sept des principales bibliothèques de recherche universitaires du Canada ainsi que l'ICIST, Bibliothèque et Archives Canada (BAC) et la bibliothèque du Parlement.

1.3 Le passé

Il n'existe cependant aucune organisation pour la conservation nationale des données et le Canada ne dispose pas non plus d'une stratégie ou de politiques nationales d'accès aux données. Les participants à la CNADRS ont exprimé de sérieuses préoccupations au sujet de la perte de données, tant comme bien national que comme éléments de base longitudinaux devant permettre la mesure du changement à travers le temps. Ces pertes se produisent à la suite de la dégradation des supports de stockage, de la perte de supports et de métadonnées, de l'obsolescence du logiciel et du matériel, ainsi que comme conséquence des politiques et décisions en matière de protection des renseignements personnels (p. ex., par les comités d'éthique de la recherche) et d'un manque de planification ou d'attention porté à la conservation au-delà de l'organisation impliquée ou du chercheur individuel.

1.4 Les mesures à prendre

Il faut donc agir d'urgence pour mettre fin à la dégradation et à la perte du patrimoine de recherche du pays; de telles mesures pourraient, par la même occasion, amener le Canada en tête de peloton dans ce nouveau contexte de recherche-développement.

Nous recommandons la création d'un groupe de travail nommé Groupe Données, chargé de préparer une stratégie de mise en œuvre nationale exhaustive et d'élaborer un projet pilote devant montrer la valeur et les répercussions de l'accès aux données de recherche par de nombreuses personnes dans de nombreuses disciplines. Lorsqu'une telle stratégie nationale aura recueilli un vaste soutien et obtenu les engagements de financement voulus, nous proposons la création d'une infrastructure nationale spécialisée, provisoirement identifiée comme Données Canada, dont l'objet sera d'assumer la direction d'ensemble de l'élaboration et de l'exécution d'un plan stratégique. Le plan engloberait et dépasserait vraisemblablement les recommandations formulées par la CNADRS.

1.5 L'avenir

Si Données Canada met en œuvre les recommandations de ce rapport, nous sommes persuadés que le pays sera en mesure de réaliser la vision identifiée par la CNADRS du rôle du Canada dans la recherche mondiale des années 2020. Nous envisageons que d'ici là :

Le Canada sera le centre d'un réseau mondial du savoir. Il sera devenu le pays avec lequel on souhaitera effectuer des recherches en partenariat, en raison de son système national de libre accès aux données de la recherche. Grâce à ce système et à la culture de collaboration qu'il aura engendré, la créativité et l'innovation canadiennes seront considérées les meilleures de leur catégorie à l'échelle mondiale. L'accès libre mais sécurisé à de puissantes base de données, réunies à l'échelle mondiale, aura vite fait de transformer la recherche scientifique. Les chercheurs analyseront couramment des problèmes d'une complexité précédemment difficile à imaginer , en quelques mois plutôt qu'en plusieurs décennies, ce qui aboutira au développement de connaissances et à des découvertes qui enrichiront la qualité de vie, transformeront les soins de santé, amélioreront l'égalité sociale, procureront une plus grande sécurité, élargiront les perspectives de décision en matière de politiques et de progrès sociaux, environnementaux et économiques et transformeront l'évolution et le développement des connaissances humaines.

Le Canada n'est pas le seul pays à nourrir des aspirations aussi élevées, et si nous tardons à agir, nous infligerons un tort cumulatif à son rôle potentiel de chef de file. Comme d'autres pays prennent déjà les devants, la temporisation est un choix qui porte en lui le germe de la destruction.

2 Recommandations

Ces recommandations sont regroupées en fonction de la ou des organisations qui, suivant nos prédictions, assumeraient la principale responsabilité de leur mise en œuvre. L'ordre dans lequel elles apparaissent n'implique aucune priorité ni aucune hiérarchie de mise en œuvre. À la section 8, nous fournissons de plus amples explications sur les raisons qui sous-tendent chaque recommandation. Les mesures nécessaires pour mettre en œuvre les recommandations sont liées temporellement à toutes les organisations chargées de leur exécution, et les objectifs ambitieux en matière de calendrier et de hiérarchie de mise en œuvre figurent à la section 9.

Responsabilités du Groupe Données

Recommandation 1 – Organisation

Que les organismes parrains constituent un groupe de travail (Groupe Données) chargé de préparer une stratégie de mise en œuvre nationale exhaustive. Le Groupe Données devrait comprendre des représentants des milieux et organismes suivants :

- le milieu de la recherche scientifique canadien;
- es organismes subventionnaires canadiens de la recherche (IRSC, CRSNG et CRSH);
- les fondations et fiducies de l'infrastructure de recherche du Canada (FCI, Génome Canada et les sociétés de fiducie provinciales équivalentes);
- les universités, collèges et instituts de recherche qui gèrent l'infrastructure de recherche au Canada;
- les ministères et laboratoires de recherche gouvernementaux (tant fédéraux que provinciaux) qui déterminent la politique publique en matière de recherche;
- l'ABRC;
- l'Association des universités et collèges du Canada (AUCC);
- le CNC/CODATA;
- l'ICIST;
- BAC;
- Statistique Canada;
- CANARIE Inc.;
- le Commissaire à la protection de la vie privée du Canada;
- des représentants d'organisations étudiantes et de professionnels qui participent à la planification de programmes de formation à la recherche;
- des représentants du grand public qui profiteront en fin de compte des bienfaits généraux de la recherche financée par l'État.

Le mandat du Groupe Données consisterait à orienter et superviser un petit secrétariat chargé de la mise en œuvre aux fins de :

- commander un projet pilote d'accès aux données (Projet Données) afin de faire la démonstration des concepts et des valeurs formulés dans ce rapport;

- planifier et superviser la formation d'une organisation canadienne permanente d'accès aux données (Données Canada);
- obtenir l'engagement à long terme du financement fédéral de Données Canada;
- élaborer un plan stratégique d'accès aux données (Plan d'accès aux données).

Recommandation 2 – Éducation

Que le Groupe Données, de concert avec les dirigeants des organismes parrains, commence immédiatement à sensibiliser les dirigeants politiques et institutionnels et les meneurs d'opinion publique aux facteurs suivants :

- le changement de contexte provoqué par l'accès à des ressources massives de données qui se produira dans le secteur de la recherche scientifique, dans le monde entier et au Canada;
- la nécessité pour le Canada de figurer parmi les chefs de file mondiaux de la transformation des modalités de la recherche, afin de conserver et de renforcer notre compétitivité économique et notre excellence scientifique à long terme;
- les avantages sociaux, médicaux, écologiques, environnementaux et économiques qui découleront de l'accélération du rythme de la découverte scientifique;
- les avantages éducatifs découlant de la capacité de situer l'apprentissage dans un contexte réel et de l'accès libre aux données;
- la nécessité d'une action concertée afin d'alimenter les changements culturels, juridiques, politiques ainsi que les changements sur le plan de la gestion qui seront essentiels à l'établissement de Données Canada.

Recommandation 3 – Financement

Qu'un financement suffisant soit fourni au Groupe Données afin de l'aider à mettre en œuvre un projet pilote de base de données à accès libre dès 2005. Ce projet devrait être conçu de façon à démontrer de quelle façon des bases de données, lorsqu'elles sont interreliées, peuvent créer d'importantes percées sur le plan des connaissances. Il comporterait également des solutions aux difficultés posées par des facteurs tels que les politiques, la technologie, l'infrastructure, la gestion des systèmes, la qualité des données et des métadonnées, l'intégrité et la sécurité. Ceci ne signifie pas qu'il faille démarrer un nouveau projet de recherche scientifique, mais il s'agit plutôt de compiler et d'intégrer des données à partir d'un exemple convaincant qui est déjà en cours ou terminé.

Responsabilités de Données Canada

Recommandation 4 – Participation internationale

Que Données Canada se dote d'une capacité de gestion qui lui permettra de surveiller les forums internationaux sur l'accès libre et d'y intervenir afin de protéger les intérêts canadiens et aider la communauté internationale à favoriser des accords, des normes et des politiques qui s'inscrivent à l'appui de pratiques modèles en matière d'accès, de partage et de conservation, compatibles avec les besoins du Canada.

Recommandation 5 – Éthique

Que Données Canada initie des consultations avec les commissions sur la protection de la vie privée, le Conseil national d'éthique en recherche chez l'humain, l'Association canadienne des comités d'éthique de la recherche, les bibliothécaires et archivistes de données et Statistique Canada afin de déterminer quels sont les obstacles juridiques à l'accès aux données scientifiques. De telles consultations devraient aboutir à des propositions de modification des lois sur la protection des renseignements personnels ou de leur interprétation juridique, afin de garantir que les données de grande valeur, financées par l'État et dont la confidentialité est convenablement protégée seront sauvegardées, sécurisées et rendues accessibles au moyen de mécanismes de contrôle de l'accès convenablement gérés. Par conséquent, les données pourront être utilisées à des fins de recherche autres que celles qui était originalement prévues et pour lesquelles un consentement éclairé peut ne pas avoir été demandé.

Recommandation 6 – Protection des renseignements personnels

Données Canada devrait entreprendre un examen de la Loi sur la protection des renseignements personnels et les documents électroniques ainsi que d'autres textes législatifs connexes afin de découvrir les incohérences qui empêcheraient le partage de données sur le plan international avec des pays dont les projets de recherche en partenariat et les pratiques de partage de bases de données devraient être de niveau élevé – les premières analyses devraient se concentrer sur les États-Unis et l'Union européenne. Données Canada devrait également collaborer avec les législateurs canadiens en matière de protection des renseignements personnels pour ajuster cette loi de façon à permettre un partage de données qui soit entièrement conforme entre des pays particuliers.

Recommandation 7 – Archivage

Que les bases et les ensembles de données, définis par Données Canada (p. ex., au moyen d'un comité d'examen de pairs ou d'un comité spécial ou d'une société ou association professionnelle compétente) comme étant d'importance nationale, soient déposées et mises en sécurité à BAC.

Recommandation 8 – Responsabilité

Que Données Canada constitue un groupe d'experts pour examiner les cadres juridiques canadien et internationaux relativement à la responsabilité pour les bases et les ensembles de données et lui demande de proposer un nouveau cadre juridique canadien compatible avec l'évolution des cadres juridiques internationaux. L'objectif recherché consisterait à équilibrer la responsabilité des gardiens de données et de leurs institutions avec les avantages sociaux résultant de l'accès libre à ces données, afin de protéger les gardiens contre la responsabilité découlant d'utilisations futures imprévues des données.

Recommandation 9 – Anonymisation

Qu'un groupe d'experts soit nommé et chargé d'examiner les problèmes juridiques afférents à l'anonymisation des données et aux pratiques en matière de données sécurisées qui empêcheraient l'invasion de la vie privée d'un particulier si elles étaient rendues accessibles pour d'autres

recherches. Le groupe devrait établir des limites à l'applicabilité du consentement éclairé, lorsqu'aucune identification n'est possible ou que l'identité de particuliers ou de petits groupes ne peut être obtenue par déduction. Si des occasions d'autoriser la réutilisation de données anonymisées se présentent, le groupe d'experts proposera des modifications appropriées aux pratiques juridiques et réglementaires.

Recommandation 10 – Bases de données à risque

Que Données Canada établisse un fonds afin de sauvegarder et d'améliorer l'accessibilité de bases de données actuelles de grande valeur, « à risque » et/ou critiques désignées par des groupes d'examen de pairs comme présentant une importante valeur actuelle, future ou historique.

Recommandation 11 – Critères et qualité

Que Données Canada collabore avec ses partenaires de recherche pour établir une fonction au sein de Données Canada (et de ses équivalents internationaux) permettant d'officialiser les critères d'évaluation de la qualité des données ainsi que de définir des processus permettant de mesurer la qualité et l'intégrité des données.

Responsabilités des organismes de financement

Recommandation 12 – Formation des chercheurs

Que tous les organismes qui financent la recherche scientifique fournissent un financement particulier devant permettre la formation de tous les chercheurs principaux aux pratiques exemplaires en matière de sélection et de gestion de bases de données, de la gestion des droits associés à la conservation de données, des normes concernant les métadonnées et des autres questions importantes, afin que l'accès et la conservation puissent être intégrés aux plans d'acquisition et d'entreposage de données dès le départ.

Recommandation 13 – Plan de gestion des données

Que les conseils de recherche et tous les autres organismes de financement de la recherche du secteur public ainsi que les ministères exigent que les demandes de fonds et de subvention comprennent un plan de gestion des données, et qu'ils identifient spécifiquement le financement requis pour garantir la qualité, l'intégrité, l'accessibilité et la reddition de compte. L'inclusion d'un plan bien élaboré pour l'acquisition, la gestion, l'accès et la préservation de données devrait être une condition du financement d'un projet de recherche. L'adhésion à de tels plans devrait également devenir une mesure non concurrentielle du rendement pour les projets et le portail permettant des demandes subséquentes de subvention. Les conseils devraient reconnaître ceux-ci en tant que coûts supplémentaires aux activités principales des projets de recherche.

Recommandation 14 – Ressources

Que les ministères et organismes gouvernementaux fédéraux et provinciaux qui financent la recherche scientifique établissent des affectations stables et non concurrentielles à long terme dans le budget central afin de fournir aux institutions, organismes et organisations de recherche les ressources dont elles ont besoin pour sauvegarder toutes les bases de données importantes (ayant une grande valeur historique, actuelle ou potentielle). Que le gouvernement fédéral fournisse également un financement supplémentaire et suffisant à BAC pour garantir la sauvegarde et l'archivage à long terme de tous les bases et ensembles de données importants.

Recommandation 15 – Examen par les pairs

Que les bases et ensembles de données en cours d'utilisation ou dont l'utilisation est prévue dans le cadre d'initiatives de recherche menées en partenariat, y compris leurs métadonnées, fassent l'objet d'un examen par les pairs et que cette évaluation fasse partie des métadonnées.

Recommandation 16 – Délais

Qu'en collaboration avec Données Canada, les organismes de financement et ministère établissent des limites à la durée pendant laquelle les gardiens des données pourront refuser le libre accès à leurs bases de données. Ce délai devrait être juste et raisonnable suivant les circonstances du moment. Après la période prescrite, les bases de données devront être rendues accessibles publiquement, sous réserve uniquement des contraintes imposées par la loi ou des protocoles et accords internationaux.

Responsabilités des universités et des chercheurs

Recommandation 17 – Récompenses

Que les facultés universitaires, le corps professoral et d'autres unités de recherche universitaires élargissent leur système de reconnaissance et de récompense destinés aux chercheurs afin d'y inclure l'excellence des contributions aux données scientifiques et l'élaboration d'outils pour l'amélioration de la gestion et de l'utilisation des données en tant qu'indicateurs de rendement importants.

Recommandation 18 – Création de spécialistes

Que les établissements d'enseignement postsecondaires accroissent leur inscription d'étudiants en sciences de l'information et l'enseignement de l'accès aux bases de données et de leur conservation afin de régler la pénurie de bibliothécaires, de gestionnaires, de conservateurs et d'archivistes numériques formés.

3 Objectifs

Les objectifs de la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS) sont de recommander aux principaux organismes et organisations de financement de la recherche les mesures qu'il est nécessaire de prendre pour maximiser, grâce à l'accès libre, la valeur des données sur le plan de la recherche et sur un plan économique, les bienfaits publics des données recueillies grâce à des deniers publics, ainsi que les mesures permettant de préserver les données qui présentent un intérêt historique comme bien scientifique et culturel pour la recherche actuelle et future. Les recommandations de ce rapport ont pour objet de donner naissance à des solutions applicables aux obstacles technologiques, institutionnels, culturels, juridiques, financiers et comportementaux à cet accès.

La CNADRS a été conçue pour compléter la *Consultation sur les archives nationales de données pour les sciences sociales*, effectuée par le Conseil de recherches en sciences humaines (CRSH) du Canada et les Archives nationales du Canada (appelées dorénavant Bibliothèque et Archives Canada – BAC).

La CNADRS a été chargée de recommander des mesures qui ne s'appliquent qu'aux données numériques. Nous avons exclus la consultation sur la question de l'accès libre aux résultats de la recherche et aux publications qui en découlent, bien que la publication de résultats soit souvent liée étroitement à l'accès libre aux données de la recherche scientifique. La question de l'accès libre et éventuellement gratuit aux résultats de la recherche et aux documents scientifiques est fortement controversée, mais devrait devenir le point de mire d'une consultation nationale spéciale dans un proche avenir. Ceci pourrait bientôt avoir des répercussions sur le développement de collaborations en matière de recherche, particulièrement à l'égard des initiatives en recherche médicale aux États-Unis et au Royaume-Uni.

4 Contexte

Introduction

Depuis que les gouvernements de tous les pays industrialisés se sont impliqués dans le financement de la recherche scientifique, on s'est toujours posé des questions au sujet de la valeur que le public devrait retirer de ce financement, par opposition aux intérêts commerciaux de l'industrie et aux intérêts financiers connexes des chercheurs, laboratoires de recherche et établissements universitaires. L'expansion rapide des connaissances au niveau mondial et la dépendance des nouvelles connaissances vis-à-vis les observations antérieures et les données de tiers, ainsi que les nouvelles bases de données qui sous-tendent ces nouvelles conclusions scientifiques, ont nourri le débat.

D'une part, une collectivité croissante est persuadée que le taux associé à la découverte scientifique est déterminé par l'accès aux données ou aux résultats de tiers et considère l'accès libre comme un facteur d'accélération fondamental de la diffusion des connaissances scientifiques. D'autre part, l'on craint qu'un tel accès minera l'industrie des services d'édition scientifique et compromettra la gestion et le rendement des investissements de la précieuse propriété intellectuelle dérivée de la recherche. Dans le domaine des connaissances médicales et sociales particulièrement, ce débat est davantage compliqué par les droits des patients opposés aux besoins d'observation longitudinale de l'évolution de la santé et de la société, besoins qui impliquent fréquemment des données qui ont été recueillies précédemment par des tiers à des fins différentes. Le domaine médical est également compliqué par les droits à la confidentialité personnelle et à la protection de la vie privée qui découlent de la Charte, dans le cadre des soins de santé fournis par les secteurs public et privé. Ceci crée un grave conflit avec les intérêts commerciaux de l'industrie des soins de santé et les bienfaits pour l'ensemble de la population.

De plus, de nombreux chercheurs ont remarqué que les premières données, sur lesquelles une grande partie de nos connaissances ont été bâties, ont déjà été perdues et continuent de l'être à un rythme qui va en s'accélérant, malgré les développements de la technologie de l'information (TI). En fait, la TI est responsable d'une grande partie des pertes, car la technologie de stockage a procuré un faux sentiment de sécurité contre les pertes et l'obsolescence. Par ailleurs, même si les données peuvent encore subsister ailleurs, il existe de très rares cas dans lesquels les données ont été systématiquement archivées et assorties de métadonnées pertinentes au sujet de leur applicabilité et leur création, ce qui permettrait de les rendre facilement accessibles et disponibles pour réutilisation, indépendamment du fait que ce service soit fourni gratuitement ou commercialement.

Lorsque l'on considère la perspective plus vaste de l'accès libre aux documents scientifiques décrivant les résultats, les conclusions et la propriété intellectuelle de l'analyse de la recherche, le débat est bien plus complexe encore. Il fait surgir les problèmes de gestion et de contrôle de la propriété intellectuelle, ainsi que les préoccupations entourant ces droits – ceux de l'industrie de

l'édition par opposition à ceux du public, des chercheurs et des institutions profitant de leur créativité, des institutions permettant d'utiliser les résultats de la recherche à des fins d'éducation et de recherche subséquente, etc.

Dans ce débat qui devient chaque jour de plus en plus public, on trouve des protagonistes qui sont d'avis qu'un accès plus libre, tant aux données qu'aux résultats de la recherche, serait fortement bénéfique pour l'efficacité des travaux de recherche. À l'opposé, il y a ceux qui voient dans l'accès libre une menace à un vaste ensemble de pratiques acceptées et éprouvées. Cependant, dans le secteur plus restreint de l'accès aux données, ce débat a atteint une intensité suffisante pour justifier la signature d'une *Déclaration internationale sur l'accès aux données de la recherche financée par des fonds publics*[1] par la plupart des pays industrialisés, les engageant par conséquent à un régime d'accès plus libre aux données.

La *Déclaration* a pour prémisse que les données de la recherche financée par des fonds publics devraient être ouvertement disponibles dans la mesure du possible. Le consensus est certes généralisé en ce qui concerne l'augmentation de l'accès aux données scientifiques, mais la mesure suivant laquelle les données devraient être ouvertement disponibles demeure fortement contestée. Le forum de la CNADRS était fortement en faveur que l'accès aux données de la recherche soit le plus libre et le plus abordable possible, mais il a reconnu que la mesure suivant laquelle ceci pourrait être réalisé devait évoluer avec le temps.

La *Déclaration* ne prend pas d'engagement au sujet de l'accès libre ou gratuit aux documents de recherche, sujet déjà chaudement débattu sur les scènes politique, universitaire et commerciale. Aux États-Unis, les National Institutes for Health (NIH), appuyés par le Congrès, prévoient actuellement d'instaurer une exigence stratégique en vertu de laquelle les résultats de toute la recherche médicale financée par des deniers publics doivent être placés dans des systèmes d'information publiquement accessibles (c.-à-d., PubMed Central), dans les six mois suivant leur publication initiale. Au Royaume-Uni, des défenseurs de la cause ont essayé également de faire adopter une proposition semblable, bien que le gouvernement se soit récemment prononcé contre. Entretemps, le Wellcome Trust du Royaume-Uni, organisme caritatif indépendant finançant la recherche sur la santé humaine et animale, prévoit de faire de cette divulgation une condition de toute la recherche financée par le Trust et prévoit lancer un équivalent européen du PubMed Central américain.

Les questions et les réponses entourant l'archivage de données ont été abordées par le Conseil de recherches en sciences humaines (CRSH) à l'intérieur même de son milieu de la recherche. On peut lire les recommandations formulées dans le *Rapport final – Consultation sur les archives nationales de données*[2]. L'examen de ce rapport s'applique largement à un contexte canadien plus vaste, mais il ne couvre que les questions portant sur l'infrastructure. Par conséquent, il n'aborde pas bon nombre des problèmes clés décrits dans le *Rapport final*[3] de suivi de l'OCDE et qui sont :

- la technologie, y compris l'interopérabilité et la gestion de la qualité;
- les institutions et la gestion, y compris la nécessaire diversité des modèles institutionnels, des données adaptées à la discipline et de la gestion de l'archivage;
- le financement des immobilisations et du fonctionnement, y compris la conservation à long terme;
- les problèmes juridiques et stratégiques, y compris les accords de collaboration internationaux sur les pratiques de partage, ainsi que les cadres législatifs des droits;
- les problèmes culturels et comportementaux, y compris la structure des récompenses, les

obstacles à la communication d'une discipline, la fierté de la propriété et la conformité en matière de structure des données.

Les objectifs de la CNADRS sont d'examiner tous les problèmes qui doivent être abordés afin de mettre en œuvre avec succès une infrastructure et une culture d'accès plus libre aux données et de recommander des mesures qui permettront d'accélérer l'évolution du système et de la culture, augmentant ainsi l'efficacité de la recherche financée par le gouvernement du Canada.

Diverses initiatives sont en cours dans le monde entier et une communauté de chercheurs, qui est déjà en voie d'augmentation, se familiarise avec tous les aspects de l'accès libre et de la conception et de l'exploitation en partenariat des bases de données. En voici quelques exemples :

- les collectivités médicales – PubMed Central, sa contrepartie de l'UE, la Base réseau de données interactives en biologie moléculaire (BIND) et la Base de données internationale sur le génome (Genbank);
- les collectivités astronomiques – le Digital Palomar Observatory Sky Survey (DPOSS), Hubble Deep Field – South (HDF-S) et le Centre canadien de données en astronomie (CCDA);
- les collectivités d'observation de la Terre – Structure des systèmes globaux de l'observation de la Terre (SSGOT);
- les collectivités environnementales – Système mondial d'information sur la biodiversité (SMIB).

Cependant, les infrastructures, systèmes et processus sont fortement hétérogènes, car il y a eu très peu de liens sur le plan technique et sur le plan de la gestion entre ces initiatives mondiales. Elles se sont néanmoins révélées être de grande valeur, car elles ont réuni des données qui ne pouvaient pas être créées autrement et ont permis l'utilisation de nouvelles approches scientifiques avec une plus forte résolution et fiabilité. En outre, bien qu'il ne soit pas facile de les transférer d'une discipline à l'autre ou à l'intérieur d'une même discipline, elles supportent quand même une plus grande complexité et peuvent être appliquées d'une façon plus générale tout en faisant davantage autorité.

À un autre égard, rares sont les pays et les disciplines qui ont réalisé des progrès considérables et généralisés en ce qui a trait à la mise en œuvre du partage de données. Par conséquent, il existe une occasion valable pour le Canada d'affirmer son leadership mondial en établissant des solutions physiques, opérationnelles, systémiques et stratégiques appropriées. Nous espérons et attendons que ce rapport constitue le premier pas vers l'élaboration d'une solution canadienne généralement acceptée et qui pourrait ensuite être proposée dans le monde entier.

Dans ce rapport, les termes « données »[4], « base de données »[5] et « ensemble de données »[6] sont utilisés selon un sens particulier qui est défini dans les notes à la fin de cette section.

Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics

Le 30 janvier 2004, le Canada et 33 autres pays, y compris tous les pays membres du G8, ont adopté une déclaration faisant état de leur engagement à collaborer à l'établissement d'un régime d'accès aux données numériques de recherche découlant d'un financement public. La déclaration

s'emploie à réaliser les objectifs et principes suivants :

- une ouverture qui tient compte des intérêts de l'accès libre afin d'accroître la qualité et l'efficacité de la recherche et de l'innovation ainsi que le nécessaire équilibre avec la nécessité de protéger les intérêts sociaux, scientifiques et économiques;
- la transparence qui rend les sources, la documentation (métadonnées) et les conditions d'utilisation disponibles et accessibles au plan international;
- la conformité juridique qui répond aux exigences juridiques nationales relatives à la sécurité, à la protection de la vie privée et aux secrets commerciaux nationaux;
- la responsabilité officielle qui préconise des règles couvrant le droit d'auteur, la propriété du crédit à la production, les restrictions d'utilisation, les arrangements financiers, l'éthique, les modalités d'octroi de licence et la responsabilité;
- le professionnalisme qui établit les règles pour la gestion en s'appuyant sur des normes et des valeurs professionnelles;
- la protection de la propriété intellectuelle qui fournit un accès libre aux termes de régimes juridiques différents applicables aux bases de données;
- l'interopérabilité qui répondra aux normes internationales d'utilisation;
- la qualité et la sécurité qui garantiront que de bonnes pratiques sont utilisées pour la production, le stockage et la gestion de l'accessibilité, garantissant l'authenticité, l'originalité, l'intégrité, la sécurité et la fiabilité;
- l'efficacité;
- la reddition de comptes.

Les prémisses sur lesquelles est fondée la Déclaration sont notamment les suivantes :

- l'échange international de données contribue de façon décisive au progrès de la recherche scientifique et de l'innovation;
- l'accès libre favorise le progrès scientifique et la formation des chercheurs;
- l'accès et la réutilisation optimiseront le rendement de l'investissement de deniers publics dans la collecte de données;
- la TI permet une augmentation importante de la portée et de l'ampleur des travaux de recherche financés par les deniers publics;
- l'occasion d'apporter des bienfaits importants à la société pourrait être compromise par une restriction indue de l'accès à de telles données et de leur utilisation;
- l'accès aux données de la recherche accroîtra la participation des sociétés en transition, contribuant ainsi à leur développement économique et social.

La Déclaration se termine par une invitation à l'OCDE d'élaborer des lignes directrices fondées sur le principe généralement accepté de faciliter l'accès économique aux données numériques de la recherche financée par des fonds publics. Les réponses à cette invitation seront très vraisemblablement examinées lors d'une future conférence de l'OCDE, lorsque la mise en œuvre de l'accès libre aux données sera inscrite à l'ordre du jour.

Consultation sur les archives nationales de données du CRSH

Le Résumé de la consultation du CRSH se lit comme suit :

> « En octobre 2000, le Conseil de recherches en sciences humaines et l'archiviste national du Canada ont constitué un groupe de travail comprenant des chercheurs et des archivistes et leur ont demandé d'évaluer les besoins en ce qui a trait à un système national d'accessibilité, de conservation et de gestion de données de recherche. Ayant

compilé de nombreuses preuves à l'appui d'un tel service de soutien à la création de connaissances par la communauté des chercheurs œuvrant dans le domaine des sciences humaines, le groupe de travail présente maintenant ses recommandations concernant la création d'un nouveau service national d'archivage de données de recherche. Ce service aura trois fonctions de base :

- Conserver les données de recherche compilées par les chercheurs et les organismes gouvernementaux, les entreprises de sondage et autres organismes, qui peuvent être utilisées par les chercheurs pour créer de nouvelles connaissances;
- Gérer les données déposées et en assurer la qualité; sélectionner les données qui seront conservées, élaborer et appliquer des normes relatives aux métadonnées, à l'authentification et à la sécurité et diffuser les données dans l'ensemble des technologies;
- Fournir l'accès aux données de recherche, y compris aux systèmes de sortie Web, aux services de catalogage, aux ententes avec les utilisateurs et les dépositaires afin de protéger la confidentialité et les droits de propriété intellectuelle et aux connexions aux autres dépositaires de données dans le monde entier.

De plus, le groupe de travail recommande qu'un nouveau réseau national d'archives de données de recherche intègre certaines autres fonctions, y compris organiser des cours de formation spécialisée dans les techniques de maniement des données, représenter les intérêts canadiens dans le développement de normes de données internationales, promouvoir l'échange des données comme pratique exemplaire en recherche, effectuer des recherches dans le domaine des sciences de l'information et de l'archivistique, et agir en tant qu'organisme central de coordination pour un réseau de services de données, dans les instituts de recherche canadiens.

L'information numérique compilée aux fins de la recherche joue un rôle de plus en plus important dans l'économie actuelle fondée sur le savoir. À bien des égards, les données sont le moteur de l'innovation et de notre capacité à résoudre des problèmes sociaux et économiques complexes. Bien que des milliards de dollars soient dépensés chaque année pour la collecte de données, le Canada ne possède pas l'infrastructure requise pour en assurer la conservation et l'accès par le public, ce qui limite le rendement du capital investi dans la recherche et mine la bonne gestion publique.

De nombreuses composantes de base requises pour la création des archives nationales de données de recherche existent déjà, notamment, les services de traitement des données des universités, les réseaux de transmission à haute vitesse, les directives juridiques et éthiques et les cadres de travail, les instituts partenaires potentiels, divers dépositaires de données et des initiatives relatives à un portail d'accès ainsi qu'une communauté de chercheurs qui produisent effectivement des données. L'élément manquant est un service de conservation, de coordination et de gestion des données.

Presque tous les pays développés ont reconnu le besoin d'un service national d'archives de données de recherche et certains pays ont plus d'une génération d'expérience dans l'exploitation de tels services. Le Canada peut bénéficier de cette expérience tout en développant un service d'archives de données de recherche compatible avec son contexte institutionnel et culturel unique. Nous possédons maintenant la capacité technologique et le savoir-faire requis pour créer un « système de confiance » en vue de fournir aux Canadiens et aux Canadiennes un service accessible et étendu qui permettrait aux chercheurs de localiser, de demander, d'extraire et d'utiliser les ressources en données de façon simple, transparente et économique tout en protégeant la vie privée, la confidentialité et les droits de propriété intellectuelle des intéressés. La Fondation canadienne pour l'innovation pourrait financer les coûts liés à l'élaboration de l'infrastructure de ce service. On estime que les coûts de fonctionnement annuels d'une installation et d'un réseau étendu sont d'environ 3 millions de dollars.

Le groupe de travail présente trois options pour ce qui est de la création d'un réseau national d'archives de données de recherche :

1. Créer, par le truchement de la législation fédérale, un réseau national d'archives de données de recherche en tant que version modifiée d'un organisme distinct constitué en vertu d'une loi. Cette option serait idéale pour la mise en place d'un service complet, d'un organisme fiable comprenant une installation centrale de conservation et de gestion des données et un ensemble de nœuds de services d'accès localisés dans des instituts de recherche. Le service profite pleinement de l'infrastructure de recherche existante, est stable à long terme et constitue une connexion directe pour les utilisateurs et les producteurs de données de recherche, et est apte à représenter les intérêts du Canada dans le développement des normes internationales de données.
2. Créer un réseau national d'archives de données de recherche sous la gouverne du Conseil de recherches en sciences humaines. Cette stratégie englobe les caractéristiques du premier modèle, mais ne requiert pas de

dispositions législatives. Elle prend avantage de la connexion directe et immédiate avec les chercheurs, de l'obligation reconnue de rendre compte et des structures de financement.

3. Créer un organisme de service spécial au sein des Archives nationales du Canada. Par la création d'une division autonome au sein des Archives nationales, cette option prend avantage de l'infrastructure et de l'expertise existantes en matière d'archivage. Cette stratégie n'a pas été retenue par les autres pays parce que la mission fondamentale d'un service d'archivage national des données de recherche est essentiellement différente. Néanmoins, à titre d'organisme de service spécial, le service pourrait possiblement être stable et en mesure de développer un système fiable d'accès aux données, de gestion et de conservation de celles-ci.

Dans le contexte du besoin canadien plus vaste, les recommandations du rapport du CRSH, qui se concentrent principalement sur l'infrastructure, constituent une position très utile à partir de laquelle on peut élargir le dialogue à l'ensemble de la collectivité de chercheurs et examiner les autres aspects qui sont décisifs à l'égard des engagements de la Déclaration ministérielle et du cadre établi par l'OCDE. »

---

[1] *Déclaration sur l'accès aux données de la recherche financée par des fonds publics*, datée du 30 janvier 2004 à Paris, France. Gouvernements signataires : Australie, Autriche, Belgique, Canada, Chine, République tchèque, Danemark, Finlande, France, Allemagne, Grèce, Hongrie, Islande, Irlande, Israël, Italie, Japon, Corée, Luxembourg, Mexique, Pays-Bas, Nouvelle-Zélande, Norvège, Pologne, Portugal, Fédération russe, République slovaque, République d'Afrique du Sud, Espagne, Suède, Suisse, Turquie, Royaume-Uni, États-Unis d'Amérique. Le texte de cette déclaration est donné à l'annexe 11.5.

[2] *Consultation sur les archives nationales de données – Rapport final*; daté de juin 2002. Infrastructure pour l'archivage et l'accessibilité des données de recherche du Canada » présenté par le groupe de travail – Consultation sur les archives nationales de données (CAND) au Conseil de recherches en sciences humaines du Canada et à l'archiviste national du Canada. Le texte de ce rapport est étudié à l'annexe 11.4.

[3] Rapport de l'OCDE « Promoting Access to Public Research Data for Scientific, Economic and Social Development » par le Groupe de suivi de l'OCDE sur les questions d'accès aux données de recherche financé par les fonds publics, mars 2003.

[4] Données : représentation conventionnelle d'une information (fait, notion, ordre d'exécution) sous forme (analogique ou numérique) permettant d'en faire le traitement automatique – *Petit Robert*.

[5] Base de données : ensemble de données logiquement reliées entre elles et accessibles au moyen d'un logiciel spécialisé – *Petit Robert*.

[6] Un ensemble de données est sous un sous-ensemble d'une base de données ou un ensemble de données créé en fusionnant deux ou plusieurs bases de données. Il est utilisé à des fins d'inférence ou de reconnaissance, mais est assemblé spécifiquement pour une inférence ou une reconnaissance particulière.

5 Vision

Dans son allocution d'ouverture devant le forum de la CNADRS, intitulée « La puissance de la vision en science : ouvrir l'accès aux données de la recherche scientifique canadienne », M. Arthur Carty, conseiller scientifique national auprès du premier ministre a déclaré :

> « La science est placée devant un choix pour ce qui est de la façon dont elle perçoit et gère son stock rapidement croissant de données de la recherche scientifique… La convergence scientifique et l'arrivée de fonds communs d'information donne naissance à de nouvelles formes puissantes de collaboration scientifique dans tous les coins du monde. La cyberscience transforme la façon dont la collectivité scientifique mondiale travaille et partage ses résultats intellectuels, analytiques et d'enquête… la cyberscience transforme également la valeur que nous attribuons aux données et la façon dont nous travaillons avec elle. »

Cette observation donne son sens véritable à la Vision de la CNADRS.

Notre Vision du monde de la recherche en 2020 :

Le Canada sera le centre d'un réseau mondial du savoir. Il sera devenu le pays avec lequel on souhaitera effectuer des recherches en partenariat, en raison de son système national de libre accès aux données de la recherche. Grâce à ce système et à la culture de collaboration qu'il aura engendré, la créativité et l'innovation canadiennes seront considérées les meilleures de leur catégorie à l'échelle mondiale. L'accès libre mais sécurisé à de puissantes base de données, réunies à l'échelle mondiale, aura vite fait de transformer la recherche scientifique. Les chercheurs analyseront couramment des problèmes d'une complexité précédemment difficile à imaginer, en quelques mois plutôt qu'en plusieurs décennies, ce qui aboutira au développement de connaissances et à des découvertes qui enrichiront la qualité de vie, transformeront les soins de santé, amélioreront l'égalité sociale, procureront une plus grande sécurité, élargiront les perspectives de décision en matière de politiques et de progrès sociaux, environnementaux et économiques et transformeront l'évolution et le développement des connaissances humaines.

Cette vision est ambitieuse, mais pour citer Arthur C. Clark : « Le monde a besoin de penseurs dépourvus d'inhibitions qui n'ont pas peur de faire des spéculations délirantes; il a également besoin d'ingénieurs conservateurs et têtus qui peuvent réaliser leurs rêves ». La CNADRS a réuni à la fois des penseurs dépourvus d'inhibitions et des chercheurs conservateurs et ce rapport trace une carte routière pour cette vision futuriste en définissant les étapes préliminaires audacieuses qui lanceront le Canada sur la voie menant à la prédominance dans le nouveau milieu de la recherche qui repose sur les données.

6 Occasions et répercussions

Les participants du forum ont cerné un vaste ensemble d'occasions et de répercussions qui découleraient de la mise en œuvre de l'accès libre aux données de la recherche scientifique ainsi que des solutions permettant de trouver les sources de données les meilleures, les plus pertinentes et les plus diverses pour chaque problème particulier. Par conséquent, un panorama de nouvelles orientations de la recherche et de connaissances s'est fait jour. Les occasions et les répercussions sont regroupées par sujet, afin qu'il soit plus facile de comprendre cumulativement ces répercussions. Cependant, l'énoncé global des répercussions est présenté en premier.

Énoncé des répercussions sur dix ans

En 2015, la participation du Canada en tant qu'un des chefs de file mondial du partage des données et de la recherche scientifique aura abouti à des transformations systémiques, culturelles, environnementales et économiques de notre recherche scientifique, dont les mesures sont incluses dans les énoncés suivants :

- toutes les institutions et tous les laboratoires de recherche financés par l'État disposent de processus, de procédures et d'une infrastructure pleinement opérationnels pour saisir, entreposer, réutiliser, archiver des données scientifiques et y avoir accès qui sont entièrement conformes à la Politique canadienne nationale sur l'accès aux données scientifiques et leur conservation, ainsi que de budgets stables à long terme afin de les maintenir comme instruments de pointe;
- le premier ministre cite souvent la Politique nationale du Canada sur l'accès aux données scientifiques et leur conservation et l'investissement dans sa mise en œuvre comme étant la transformation la plus importante qui a permis au Canada d'atteindre son objectif d'être le cinquième pays du monde pour ce qui est de la prédominance de la recherche;
- les entreprises canadiennes d'accès aux données, d'exploration et de conservation de données figurent parmi les chefs de file mondiaux en solutions pour la gestion de données;
- le Canada est un chef de file mondial reconnu des pratiques, processus et normes de gestion des données scientifiques, comme le montre l'adoption de systèmes canadiens de gestion des données scientifiques par la majorité des pays qui ont acquis de tels systèmes;
- les conseils subventionnaires du Canada signalent que tous les titulaires de subventions de recherche participent énergiquement à la création, à l'expansion, à la constitution et au maintien de bases et d'ensembles de données qui sont le fondement des connaissances scientifiques du Canada, ce qui a pour résultat que la productivité de la recherche scientifique par habitant du Canada figure parmi les trois plus élevées du monde;
- les laboratoires de recherche du secteur public et les universités du Canada signalent un nombre plus élevé que jamais de collaborations de recherche internationales, et une proportion plus importante de collaborations internationales que n'importe quel autre pays;
- presque tous les documents de recherche universitaires et industriels canadiens citent la réutilisation de données scientifiques de sources de données à accès libre canadiennes et internationales;

- toutes les données scientifiques d'importance nationale examinées par les pairs ont été sauvegardées dans les archives de données nationales et un financement durable est enchâssé dans le budget central à long terme du gouvernement fédéral, et son maintien recueille le consensus politique de tous les partis;
- le rendement des chercheurs de laboratoires universitaires et financés par l'État comprend l'évaluation, dans leurs institutions et par les conseils subventionnaires, de leurs contributions aux bases de données à accès libre nationales et internationales au même titre que toutes les autres contributions en matière de recherche.

Nouvelle science

Les participants du forum de la CNADRS ont prévu que de nombreux nouveaux secteurs scientifiques émaneront de l'accès à des sources de données mondiales actuelles et nouvelles et à de nouveaux croisements de celles-ci. Ils découleront principalement de nouveaux croisements de disciplines et de données qui :

- donneront naissance à de nouvelles méthodes de recherche reposant sur l'entrecroisement de données qui ne pouvaient pas être réunies en 2005;
- amèneront une réduction de la complexité qui n'est réalisable que grâce à l'intégration et à l'exploration de données interdisciplinaires;
- permettront l'étude de données transitoires et de celles qui sont saisies en permanence;
- créeront des relations imprévues et inconnues entre des données qui sont trouvées par hasard lorsque d'autres travaux de recherche sont entrepris.

Les participants du forum sont persuadés que d'importantes percées scientifiques se produiront par suite de l'existence de bases et d'ensembles de données accessibles d'une taille et d'une portée qu'il était précédemment impossible de construire, en raison des contraintes imposées par la taille des équipes, les frontières des disciplines, les capacités de gestion de bases de données et les normes associées aux métadonnées. L'incidence de l'accès libre a transformé la recherche des connaissances simplement grâce à l'échelle et à la qualité des données accessibles aux chercheurs.

Exemples des progrès et des défis de 2000 à 2005

Accélération de la recherche reposant sur les données

Les communautés de chercheurs en **génétique et en génomique** ont déjà mis au point un modèle éprouvé d'accès et de partage de données. Les bases de données sur l'ADN et les protéines ont déjà révolutionné la façon dont on travaille en biologie. Ce modèle est le fondement de l'accélération de la recherche en bioinformatique. La mise au point continue de processus dynamiques afin d'identifier automatiquement les données brutes dérivées directement des systèmes de séquençage d'ADN accélèrera encore la recherche en bioinformatique. À la suite de l'accumulation d'ensembles très vastes de données, le rassemblement et l'analyse d'expériences multiples qui sont effectuées au plan international devient réalisable pour la première fois. Des subtilités qui autrement seraient restées indétectables ou seraient passées inaperçues seront dorénavant évidentes et mèneront à l'émergence de nouvelles idées. Les progrès réalisés dans cette discipline ne sont que le début d'une nouvelle vague d'occasions. Les interactions et les découvertes qui deviennent possibles lorsque d'énormes bases de données – médicales, environnementales, sociales, économiques et démographiques – peuvent être coanalysées, permettront aux chercheurs de commencer à comprendre véritablement les

nombreux facteurs qui contribuent à améliorer la compréhension universelle, y compris de la vie et de la santé.

Difficultés posées par les contraintes juridiques

Dans le **milieu de la santé,** l'accès aux données de la recherche et leur réutilisation ont été fortement restreints en raison de problèmes de protection de la vie privée et de la Loi sur la protection des renseignements personnels et les documents électroniques. Les nouvelles méthodes permettant de garantir la confidentialité, notamment la gestion de l'accès sécurisé en vertu de protocoles rigoureusement appliqués qui garantissent l'application correcte du consentement éclairé et le chiffrement efficace des données, élimineront la nécessité des pratiques de destruction des données qui prévalent aujourd'hui dans les milieux de recherche en santé. Ces méthodes permettent le rassemblement de grands ensembles de données de recherche qui, quant à eux, permettent d'élucider les causes, les modalités de traitement et les traitements de substitution (incluant les combinaisons de traitement), les pronostics et les résultats de problèmes communs et moins communs de santé et permettent la production d'études à long terme qu'il aurait été impossible ou statistiquement peu valide de réaliser en raison de la perte de données ou de la faible taille de l'échantillon disponible pour les pratiques actuelles. Lorsque de tels changements dans les infrastructures de données seront en vigueur, les chercheurs auront la possibilité de combiner leurs nouvelles données à de nombreuses autres bases de données afin que des facteurs nouveaux puissent émerger.

Nouvelles connaissances découlant de bases de données pluridisciplinaires reliées

Dans le **milieu de la recherche environnementale,** il existe déjà de nombreuses bases de données nationales et internationales dont bon nombre sont reliées. Notre compréhension des interactions complexes des océans, de la terre et de l'atmosphère, ainsi que les actions des êtres humains, est de plus en plus alimentée par des bases de données mondiales dont les éléments sont produits par des capteurs dans ces secteurs et par des données en provenance d'imagerie satellitaire à large spectre. De tels liens feront naître une compréhension totalement nouvelle de notre environnement et des changements qui s'y déroulent. Les efforts déployés pour identifier de tels changements deviendront pour la première fois mesurables et observables. De plus, ces nouvelles connaissances créeront des occasions commerciales pour de nouveaux produits et services pour les entreprises canadiennes.

Meilleures connaissances scientifiques

L'accès libre aux données de la recherche permettra d'améliorer les connaissances scientifiques. Par le passé, les données sur lesquelles reposait la recherche scientifique ne présentaient souvent pas la précision, l'exactitude et la quantité nécessaires pour valider de nombreuses thèses de recherche avec un degré de confiance élevé. Par conséquent, les connaissances et les hypothèses engendrées sont toujours restées ouvertes à une réévaluation et/ou une réinterprétation. Une grande partie de nos connaissances étaient fondées sur des hypothèses au sujet de la précision minimale nécessaire et de la plus faible fréquence d'observation des données, ce qui a abouti à des résultats incohérents et à des connaissances scientifiques peu fiables. Ceci était habituellement provoqué par la durée et le coût de production et de validation des données. Qui plus est, dans de telles circonstances, à moins que les données originales aient été conservées sous une forme qui peut être réutilisée, toutes les réévaluations et réinterprétations doivent être effectuées uniquement sur des données nouvellement créées ou saisies, et les bases de données historiques ne peuvent donc être ni vérifiées ni analysées de nouveau.

De meilleures connaissances scientifiques résulteront de la capacité d'observer, de réévaluer et d'analyser de nouveau les données originales et de les examiner dans le contexte des nouvelles connaissances et/ou des nouvelles données. Elles proviendront également d'ensembles mondiaux de données beaucoup plus importants qui seront créés lorsque des équipes internationales collaboreront. L'augmentation de la fréquence d'observation et l'amélioration concomitante de la validité des données réduiront les erreurs et encourageront une interprétation et une compréhension plus exhaustives des écarts de données observés. L'accès libre, assorti d'un archivage systématique et de métadonnées améliorées et normalisées permettra également d'effectuer des analyses longitudinales de qualité supérieure, une meilleure réplication des expériences et un examen mieux informé par les pairs.

Exemples :
Il existe quelques exemples dignes de mention de meilleures connaissances scientifiques qui ont découlé du besoin mondial de fournir un accès libre aux données scientifiques portant sur des problèmes particuliers, notamment :

- les principes de Bromley de 1991 qui instauraient un échange intégral et libre des données de la recherche sur le changement de l'environnement mondial et qui ont permis la création du SSEOT;
- les principes des Bermudes de 1996 sur la divulgation de données sur la séquence du génome humain qui ont abouti à la création de GenBank.

Dans ces cas et dans de nombreux autres, les protocoles de libre accès ont entraîné des progrès notables des connaissances qui ont surpris même les chercheurs. À l'heure actuelle, l'exemple le plus significatif est sans doute la cartographie du génome humain.

Leadership en innovation

L'accès libre aux données de la recherche scientifique stimulera un changement en série de la culture de la recherche scientifique, et chaque évolution culturelle encouragera à son tour une nouvelle expansion de l'accessibilité. Au fur et à mesure que la capacité de travailler par delà les frontières des disciplines augmentera et révèlera la valeur énorme des nouvelles connaissances qui en résulteront, le taux d'innovation augmentera. Le Canada ne sera certes pas le seul pays à connaître un tel essor de l'innovation, mais la stratégie intégrée que nous recommandons et l'adoption prompte de l'accès libre en tant que priorité nationale garantiront qu'il occupera une position de leader parmi les pays axés sur le développement de la recherche.

Le forum de la CNADRS a notamment relevé que les obstacles interdisciplinaires qui ont empêché le développement de nouveaux secteurs commerciaux et la création d'occasions d'affaires disparaîtront lorsque l'accès libre sera instauré. L'un des principaux secteurs à exploiter est la prévention des maladies. L'analyse de la prédisposition génétique, sociale et environnementale à des maladies particulières permettra l'expansion très rapide du « secteur bien-être » de l'économie. Cette percée permettra le traitement préventif dans le cas de prédisposition à la maladie et éliminera son apparition ou atténuera significativement ses effets. D'autres secteurs seront stimulés de la même façon par l'accès libre, notamment :

- l'industrie des logiciels qui élaborera des systèmes et outils améliorés de recherche, d'exploitation et de gestion des données;

- les industries qui fabriquent des détecteurs de saisie des données pour la recherche scientifique et pour ses utilisations à des fins commerciales, industrielles, environnementales, sociales, de soins de santé et de bien-être.

Politique et stratégie supérieures

La capacité de recueillir des données, d'y avoir accès et de les combiner à partir d'un réseau de bases de données, aussi bien actuelle qu'archivée, auxquelles on fait confiance, permet aux conseillers en élaboration de politiques et aux planificateurs stratégiques d'examiner les effets des politiques et stratégies de substitution dans un contexte beaucoup plus vaste qu'il n'est possible de le faire aujourd'hui.

Ainsi, les répercussions des politiques en matière de transport sur les coûts de santé à long terme et la productivité du milieu de travail deviennent possibles grâce à une meilleure compréhension de modèles de pollution et des effets de la pollution sur la santé, ainsi que des coûts de santé qui en résultent et des effets connexes sur la productivité. La capacité de déduire une compréhension plus vaste des répercussions de politiques et stratégies de substitution grâce à une modélisation reposant sur l'accès libre présente un potentiel énorme de retombées pour la société. Ces dividendes ne peuvent être produites ni même calculées aujourd'hui parce que la combinaison des données essentielles dépasse de loin la capacité des analystes de politique, des planificateurs stratégiques, voire des chercheurs eux-mêmes.

De nombreux autres exemples peuvent être présentés, notamment les efforts mondiaux déployés pour créer et intégrer dans l'avenir des politiques et stratégies environnementales internationales et nationales. Par le passé, de telles initiatives n'ont remporté qu'un succès mitigé, parce que les données nécessaires n'étaient pas facilement accessibles ni d'une grande cohérence :

- La dynamique du changement environnemental, bien qu'elle soit fondée sur des données environnementales actuelles, n'est pas toujours convenablement comprise et l'impact de l'activité humaine est sujet à interprétation. Par conséquent, le public nourrit un scepticisme justifiable au sujet de son importance et de sa validité – comme le font de nombreux chercheurs. Ceci a abouti à une diversité de points de vue des pouvoirs publics au sujet de l'ampleur du défi à relever et de son échéancier. Lorsque l'on combinera des données et des ressources internationales et nationales et qu'on les reliera à d'autres données, notamment les bases de données démographiques, sur l'utilisation des sols et autres facteurs économiques, les chercheurs et les décideurs seront en mesure d'augmenter leur compréhension, de réduire l'incertitude interprétative et de renforcer les arguments pour l'obtention de mesures mieux définies.
- Les répercussions sur l'économie, la santé, la population et la société du changement environnemental continuent d'être spéculatives et douteuses, parce que les données utilisées pour ces projections sont rarement bien adaptées aux besoins du grand public et difficiles d'accès; en outre, les bases de données en provenance de pays étrangers sont rarement comparables avec leur équivalent canadien. Si les Canadiens et d'autres surmontent ces difficultés, ils pourront parvenir à des conclusions qui comportent moins d'erreurs inexpliquées, qui jouiront d'une plus grande confiance et qui attireront un plus fort soutien politique et public pour toute politique et stratégie correctrice qui pourra s'avérer nécessaire.

Recherche plus efficace

La recherche scientifique est en règle générale entravée par ce qui suit :

- des difficultés d'acquisition ou de construction de bases de données avec la précision, l'exactitude et à l'échelle voulues;
- la participation d'équipes interdisciplinaires qui apportent une expertise essentielle à la recherche; et
- l'assemblage de ressources informatiques qui ont la puissance et la rapidité requises pour modéliser des problèmes difficiles.

L'ouverture de l'accès aux données scientifiques abordera progressivement les problèmes suivants : la difficulté d'acquisition; l'accès à l'expertise requise grâce à la normalisation des métadonnées et à la création de partenariats avec des chercheurs dans d'autres disciplines; la capacité croissante d'utiliser les ressources en réseau pour régler le problème de la puissance de calcul.

Ces changements auront pour effet :

- de réduire les efforts et le temps requis pour acquérir des données de la précision, de la qualité et à l'échelle voulues;
- de permettre l'accès aux données et d'aider à réunir les chercheurs possédant des connaissances et une expertise précieuse du domaine;
- de permettre l'accès aux outils et ressources informatiques nécessaires pour répondre aux exigences des analyses complexes.

Ces nouvelles capacités accéléreront la recherche tout en réduisant l'investissement nécessaire à l'élaboration de nouvelles connaissances et au règlement de problèmes d'une bien plus grande complexité. L'effort engagé pour acquérir des données dans le cadre de projets distincts ayant livré des résultats négatifs deviendra également accessible à d'autres équipes de recherche, engendrera une valeur ajoutée à partir des efforts de recherche qui n'ont pas été couronnés de succès et créera des occasions pour le développement de nouvelles connaissances.

L'effort engagé pour bâtir de grandes bases de données polyvalentes et les processus de récompense pour ceux qui sont impliqués seront mieux compris et permettront la prise de décisions de qualité supérieure pour optimiser l'investissement entre :

- l'acquisition de données, leur accessibilité et leur archivage;
- l'énoncé de théories, la vérification, la création de connaissances et les avantages qui en découlent.

Amélioration de l'éducation

Le processus de l'enseignement scientifique a traditionnellement été fondé sur l'apprentissage de la discipline et de la théorie, et ce n'est qu'aux cycles supérieurs que l'observation et la cueillette de données scientifiques ont été introduites comme composante fondamentale. Même à ce niveau, l'approche traditionnelle a été limitée à des données propres à la discipline. L'ouverture à l'accès à des données dans toutes les disciplines encouragera les enseignants et professeurs à réévaluer leur enseignement. L'enseignement et la recherche seront modifiés par suite de l'accès libre qui permettra d'utiliser les données en développement et celles archivées d'une façon qui complètera et renforcera les connaissances que les enseignants ont pour mission de transmettre. En outre,

l'accès libre permettra de mettre en œuvre des projets de recherche avec une plus grande facilité au niveau du premier cycle et même de l'enseignement gradué. Les jeunes chercheurs pourront utiliser de véritables données, ce qui leur permettra de construire des réalités concrètes à partir des connaissances qui leur sont enseignées. Les étudiants seront en mesure d'élargir leurs horizons au-delà de ceux du programme d'enseignement obligatoire et même d'entreprendre indépendamment des recherches.

L'enseignement étant fondé sur des données scientifiques accessibles, les nouveaux processus d'enseignement permettront aux étudiants d'acquérir une meilleure compréhension de la puissance des bases de données en tant qu'outils de découverte et de résolution de problèmes. Ils pourront également en savoir davantage sur la capacité de gérer de façon plus efficace des données et de produire des résultats précieux et utiles. Par conséquent, non seulement la valeur des étudiants diplômés sera-t-elle accrue pour la société, mais grâce à leurs nouvelles connaissances et à leur respect pour les données scientifiques, ils contribueront à transmettre ce respect aux générations précédentes et suivantes.

La découverte scientifique provenant d'étudiants de premier cycle et d'étudiants gradués est relativement rare, mais pas inconnue. Grâce à l'accès libre, il sera certes plus fréquent que des étudiants exceptionnels auront la capacité de faire des découvertes scientifiques, ce qui augmentera la croissance générale des connaissances et leur fournira éventuellement une occasion d'expérience exceptionnelle qui pourra inspirer des carrières en recherche scientifique.

7 Défis de l'accès libre

Les participants à la CNADRS ont décrit un nombre important de facteurs qui entravent l'acceptation et la mise en œuvre généralisée d'un système canadien de libre accès et de conservation. Ces facteurs vont des enjeux et priorités politiques à la culture canadienne en matière de recherche et de systèmes de valorisation. Faute de mesures pour s'attaquer à ces facteurs inhibiteurs, la capacité d'instaurer l'accès libre comme moyen permettant la transformation de la recherche canadienne sera compromise, et le Canada prendra du retard sur d'autres économies avancées. Qui plus est, toute approche systématique ultérieure sera compliquée par l'existence de silos à libre accès découlant d'initiatives individuelles, de mesures indépendantes des conseils subventionnaires, d'approches disciplinaires ou institutionnelles distinctes ou de grands projets internationaux qui ont adopté des normes, des processus et des infrastructures différents. Il sera fortement perturbant, coûteux, fastidieux et bien plus exigeant en termes de ressources de restructurer un grand nombre d'approches différentes que d'agir de la bonne façon dès le départ dans le cadre d'une stratégie nationale.

Priorité des besoin

Les universités, les organismes subventionnaires et les laboratoires de recherche ont d'autres enjeux hautement prioritaires (p. ex., le financement, les programmes d'enseignement, l'expansion communautaire, l'infrastructure vieillissante). Pour que le changement se produise, l'importance du libre accès et la conservation des données doivent devenir une priorité importante dans l'ensemble des milieux de la recherche et des milieux universitaires.

Au niveau politique, il y a peu de compréhension publique concernant cet enjeu, à l'exception de la préoccupation nationale croissante au sujet de la protection des données personnelles et de leur éventuel mauvais usage, qui est perçu en soi comme l'antithèse du libre accès. Par ailleurs, il n'est pas encore clair que l'accès libre aux données de la recherche dans les universités relève des gouvernements fédéral ou provinciaux.

Les mécanismes de financement pour la construction, l'alimentation, l'entretien et la conservation de bases de données scientifiques n'existe que rarement à l'extérieur des activités de Statistique Canada et de l'ICIST. La FCI a fourni et continuera de fournir des fonds pour la capitalisation des bases de données destinées à des initiatives de recherche stratégiques, mais elle ne peut financer l'acquisition de données, leur maintien ou leur conservation. Il existe des cas exceptionnels dans lesquels les communautés de chercheurs n'ont elles-mêmes pas eu d'autres choix que d'engager des ressources pour bâtir et maintenir les grandes bases de données essentielles à la complexité des questions de recherche qu'elles sont appelées à résoudre (p. ex., astronomie, génomique, protéomique, environnement, etc.). Cependant, la conservation n'a été abordée que par les bibliothèques de recherche sous la direction de l'ABRC, et peu de financement a été mis à la disposition des gens concernés par ces besoins de conservation.

Les règles et les critères de sélection existants en matière de demande de subvention du CRSNG et de l'IRSC excluent en règle générale le financement de projets dont le but est de maintenir la capacité de recherche, notamment des bases de données, ou pour créer de telles bases de données, à moins que la technologie sous-jacente soit considérée une initiative de recherche en soi. Les récentes règles du CRSH permettent certes un tel financement, mais peu de subventions ont effectivement permis l'obtention de fonds à cette fin.

Champions du changement

Le fait que les deux conseils subventionnaires, la FCI et Industrie Canada ainsi que le CNRC sont des parrains de la CNADRS témoigne d'une inquiétude croissante à l'égard de la nécessité de favoriser le libre accès aux données de la recherche scientifique. La ratification par le gouvernement du Canada de la Déclaration sur l'accès aux données de la recherche de l'OCDE est un autre indicateur encourageant. Nous pensons cependant que des études, des débats et des signatures ne suffisent pas. Le moment est venu pour que des personnalités influentes au sein du milieu canadien de la recherche, de l'appareil gouvernemental, de la Chambre des communes et du Sénat reprennent le flambeau afin que des mesures soient prises promptement.

Culture

Les ministères et organismes fédéraux ont en règle générale établi comme pratique d'exiger des droits pour l'accès aux données afin de recouvrer les coûts de l'établissement et du maintien des données dont ils ont le contrôle. De telles pratiques risquent de mettre les données hors de la portée de nombreux chercheurs qui pourraient autrement ajouter de la valeur et faire de nouvelles découvertes grâce à cet accès. Dans certains cas, cet obstacle risque d'empêcher que l'on poursuive une recherche particulière, ce qui peut avoir les conséquences suivantes :

- des préjudices causés à la recherche canadienne et à sa compétitivité mondiale;
- l'absence d'une utilisation plus vaste par l'industrie et le commerce;
- la perte de valeur perçue qui entraîne en fin de compte la décision d'interrompre l'investissement de soutien et la perte de données qui s'ensuit.

La recherche canadienne est en grande partie animée par la curiosité de chercheurs individuels et se caractérise souvent par la prépodérance de l'intérêt personnel et une approche qui valorise la compétitivité plutôt que la collaboration. Par conséquent, un système de récompenses fondé presque exclusivement sur la reconnaissance individuelle s'est imposé. Lorsque les chercheurs s'emploient à atteindre des objectifs personnels, à acquérir des récompenses institutionnelles et à gagner la reconnaissance de leurs pairs, la culture de collaboration et de contribution en équipe au bien public est régulièrement reléguée en seconde place. Par conséquent, la motivation qui pousse les chercheurs à contribuer à des données librement accessibles devient une faible priorité pour la plupart d'entre eux. Dans une telle culture, les données deviennent principalement des moyens permettant d'obtenir une reconnaissance individuelle. À l'opposé, le cadre proposé de libre accès exige des contributions collégiales à des données de plus grande valeur afin de susciter une motivation nourrie par une reconnaissance équivalente à la reconnaissance par les pairs et la récompense habituellement accordée pour le développement de nouvelles connaissances. Le libre accès et la conservation exigent une modification du système de reconnaissance et de récompense afin de répondre aux aspirations professionnelles, particulièrement celles des jeunes chercheurs, grâce à l'excellence de la production ou de la collecte de données de qualité.

Formation

Peu de chercheurs ont reçu une formation particulière sur l'élaboration et la conservation de bases de données, et bon nombre hésitent à assumer une responsabilité pour la gestion de données qui dépassent leurs intérêts immédiats. On trouve également peu d'expertise dans les institutions et organismes de recherche scientifique consacrés à la gestion des données, qui réponde aux exigences en matière d'accessibilité, de sécurité et de conservation. De toute évidence, il devient urgent de développer cette capacité, car il faut donner une formation aux chercheurs scientifiques, et il est impératif qu'une expertise soit disponible dans les milieux de recherche pour compléter cette formation.

Normes et processus

Le libre accès interdisciplinaire généralisé aux données scientifiques exige l'adhésion à des normes et, par conséquent, de telles normes devront être mises en œuvre indépendamment d'un quelconque domaine. Or, la plupart des bases de données existantes utilisent au contraire des approches et architectures propres à une discipline qui permettent aux chercheurs spécialistes de ce domaine immédiat de les utiliser efficacement dans la recherche de nouvelles connaissances, mais qui constituent un obstacle potentiel pour tout autre utilisateur en puissance. De telles bases de données pourraient certes devenir librement accessibles, mais leur valeur pour d'autres communautés de chercheurs serait vraisemblablement fortement diminuée. Cependant, le remaniement de l'architecture, la restructuration et le remaniement des bases de données existantes et la refonte de leur interface avec des outils d'analyse, d'exploitation et d'application – sans mentionner leurs métadonnées – exigeront des efforts exceptionnels de conversion pour les communautés établies qui ont déjà élaboré et maintiennent de grandes bases de données. Ce problème peut être atténué par l'adoption, dans le meilleur délai possible, d'une stratégie de données nationale et la conformité à celle-ci grâce à la migration de bases et d'ensembles de données existants.

Dans de nombreux domaines de la recherche médicale et sociale, la législation en matière de protection de la vie privée, notamment les règles d'éthique élaborées pour empêcher d'y contrevenir, a inculqué la pratique de détruire les données après qu'elles aient été utilisées pour les fins auxquelles elles étaient destinées. L'impact de cette législation est renforcé davantage par les préoccupations entourant la confidentialité de données tels que l'ethnicité, le sexe, le choix de mode de vie, la richesse, etc., et ce, dans tous les domaines scientifiques et dans la recherche qui couvre tout aspect concernant la sécurité nationale et la propriété intellectuelle. La sécurité des bases de données, leur gestion, les contrôles et les autorisations d'accès dans la plupart des situations exigent un niveau de connaissances et d'expertise que l'on trouve habituellement uniquement dans les organismes de sécurité. Il faudra élaborer des processus et éventuellement adopter des lois pour que les comités d'éthique et les organismes de sécurité puissent accorder la permission pour la conservation et finalement l'archivage de telles bases de données si elles respectent des normes spécifiques d'anonymat.

Expertise en matière d'archivage

Il n'existe pas de nos jours assez d'archivistes formés au Canada (eu égard à la demande croissante) possédant des connaissances sur le catalogage de données, sur les normes et processus de métadonnées, sur la gestion de la conservation et l'évaluation de la valeur des données. Il n'est pas

seulement nécessaire de former des chercheurs en architecture, gestion, normes et pratiques exemplaires en matière de bases de données, mais il faut également disposer de l'expertise connexe pour guider l'élaboration de bases de données perfectionnées et aborder les problèmes d'accès aux bases de données, de leur réutilisation et de leur conservation.

Les archivistes de données sont actuellement reconnus comme des partenaires précieux de la recherche et comme des experts-conseils dans un nombre sans cesse croissant de projets qui dépendent de bases et d'ensembles de données importants. Cependant, ils ne sont pas encore considérés globalement comme étant essentiels à la recherche dans les instituts de recherche et demeurent vulnérables aux pressions budgétaires, surtout lorsque de tels « frais généraux de bibliothèque » exigent des augmentations budgétaires. Le nombre d'archivistes de données au sein des instituts a augmenté certes, mais pas proportionnellement au taux de la demande. Le passage à des modalités de recherche exigeant de grandes quantités de données s'accélérera lorsque les chercheurs canadiens adopteront rapidement ce nouveau contexte de recherche tout en essayant de demeurer concurrentiels avec leurs pairs. Ceci fera ressortir la pénurie d'archivistes et de gestionnaires de données et poussera à la fois à recruter ces ressources spécialisées et à dégager les budgets institutionnels nécessaires pour les employer.

Responsabilités, systèmes et outils

Sur un plan conceptuel, les responsabilités concernant la propriété de données acquises pour un projet unique sont simples, bien qu'elles ne soient pas suffisamment définies par la loi. Parmi les divers régimes administratifs pour la protection et la gestion de la propriété intellectuelle (PI), les bases de données disposent de leurs propres lois en matière de droit dans la plupart des pays. Ces lois ressemblent généralement à la loi sur le droit d'auteur, mais appartiennent à leur propre domaine. La nécessité d'établir un meilleur régime de PI a certes été déterminée, mais les progrès ont été plutôt lents à l'Organisation mondiale de la propriété intellectuelle (OMPI) et dans ses bureaux nationaux affiliés. À l'heure actuelle, aucun délai n'a été fixé pour aborder les problèmes clés en matière de PI de données, que ce soit au Comité permanent des droits d'auteur et des droits connexes ou lors de conférences internationales de l'OMPI. Lors de ces débats, les questions relatives aux grandes bases de données à accès libre constitueront un défi entièrement nouveau et il pourrait, par conséquent, falloir un certain temps pour que la collectivité internationale parvienne à un consensus. Entre-temps, dans le contexte canadien, la propriété, la garde et le contrôle des données sont des concepts nouveaux pour une majorité de chercheurs. La CNADRS est parvenue à la conclusion que dans un avenir immédiat, la seule solution logique pour la conservation et la garde de bases de données consiste à ce que les contributeurs de données et les conservateurs de bases de données assument une responsabilité conjointe. Ceci pourra être intégré au cadre du protocole international lorsque des régimes nationaux et internationaux pertinents de PI des données seront établis.

À un moment donné, soit à la fin de l'utilisation active d'une quelconque base de données scientifiques ou chaque fois que le conservateur des données n'est plus disposé ou capable d'en exercer la garde et le contrôle, un mécanisme doit s'enclencher pour transférer ses responsabilités à des successeurs valables, soit d'autres chercheurs compétents, des bibliothèques et archives institutionnelles ou des organismes ou archives provinciaux ou nationaux. En outre, dans les cas où des bases de données sont réputées constituer un bien national, l'archivage sécurisé de telles bases de données doit être effectué par une entité digne de foi et de confiance, par exemple, Bibliothèque et Archives Canada afin d'éliminer le risque de contamination, de destruction ou d'autres pertes.

Pour rendre les bases de données véritablement accessibles, les descriptions de métadonnées, les plateformes technologiques, les outils d'accès et d'exploration et les systèmes de gestion et de maintenance doivent être stables et d'usage commun. Ils doivent également pouvoir bénéficier du soutien pendant toute la période d'utilité des données ou être déplacés dans un milieu de rechange lorsque la stabilité, l'utilisation régulière ou le soutien ne peuvent plus être garantis. Il faut appliquer une protection semblable pour l'accès aux archives et elle devrait comprendre l'obsolescence et/ou les pannes des supports de stockage et des lecteurs de stockage, ainsi que les programmes d'accès et de maintenance.

Autres défis et occasions

Les participants de la CNADRS ont également souligné des difficultés dans plusieurs autres secteurs :

Il existe des bassins importants de données existantes qui ne sont pas utilisées à l'heure actuelle, mais qui présentent soit une importante valeur potentielle pour des analyses historiques et des études longitudinales (pour déterminer des évolutions historiques et comme base pour la recherche future) ou qui sont réputées avoir une signification historique nationale. Comme ces collections de données sont vastes, on s'inquiète de l'ampleur de la difficulté du recouvrement et de l'archivage d'objets historiques aussi importants. Des complications surgissent en raison de l'incertitude au sujet de la propriété, de l'inaccessibilité des connaissances requises pour créer les métadonnées nécessaires, de l'illisibilité du support de stockage des données ou de l'inexactitude de l'information concernant l'intégralité, ainsi que de la perte d'information au sujet de leur emplacement. Comme des données historiques précieuses ont déjà disparu et continuent de le faire à un taux toujours croissant, il est urgent de commencer le recouvrement et l'archivage. Ceci est davantage compliqué par la pénurie d'archivistes relevée ci-dessus.

D'un point de vue juridique, la responsabilité du propriétaire et le contrôle de la PI et la contrefaçon sont également reconnus comme des obstacles à la liberté d'accès. Il ne semble pas exister de précédents juridiques indiquant la responsabilité qu'un propriétaire ou gardien de données assume lorsque des données sont utilisées par d'autres chercheurs à des fins qui échappent au contrôle du gardien. En règle générale, on s'accorde à penser que le droit d'auteur des bases de données a rarement été enregistré, mais il est manifeste que les propriétaires ou gardiens de bases et d'ensembles de données précieuses recourront davantage au processus de protection de la PI. Le contrôle grâce au droit d'auteur et le libre accès sont compatibles, en ce sens que le libre accès n'empêche pas d'accorder ou de refuser la permission d'accès dans certaines circonstances, ni d'intenter des poursuites en cas de contrefaçon. Cependant, ils ne sont pas interchangeables, car ils font appels à des modes de pensée différents : accès exclusif uniquement pour ceux qui ont la permission par opposition à la liberté d'accès pour tous. Les problèmes de responsabilité, de contrôle de la PI et de la contrefaçon doivent être réglés pour éliminer ce facteur inhibiteur au libre accès aux données de la recherche scientifique.

8 Recommandations émanant de la consultation

Les recommandations de la CNADRS résultent directement du forum. Il convient de noter qu'elles concordent étroitement avec les recommandations du rapport « *Scientific Data and Information* »[1] du Conseil international des unions scientifiques (CIUS), publié en décembre 2004 par le Assessment Panel du Committee on Science Planning and Review.

Responsabilités du Groupe Données

8.1 Démarrage

La CNADRS est parvenue à la conclusion que le libre accès aux données de la recherche scientifique permettra, parmi d'autres avantages :

- de transformer le processus même de la découverte scientifique en procurant la capacité d'accéder rapidement à des ensembles de données de plus grande envergure, plus rigoureux, complets et divers qui ont été ou seront constitués grâce à un financement public;
- d'accélérer le rythme du développement des connaissances grâce à la réutilisation de données et à l'interconnexion de divers ensembles de données;
- d'étudier des systèmes et des interactions de systèmes bien plus complexes;
- de créer des occasions pour une augmentation importante de la collaboration nationale et internationale en recherche;
- d'aboutir à une myriade d'avantages économiques, environnementaux, écologiques et sociaux dans tous les domaines de la science, dont bon nombre sont inattendus, par suite de nouvelles interrelations qu'il reste encore à établir.

L'initiative CNADRS devait produire, à l'intention de ses parrains, des recommandations à l'égard de la justification d'un système canadien d'accès aux données scientifiques et de leur sauvegarde et des mesures pour sa création. La première recommandation de la CNADRS propose de créer une structure organisationnelle pour commencer la mise en œuvre de cette initiative :

Recommandation 1 – Organisation

Que les parrains constituent un groupe de travail (Groupe Données) chargé de préparer une stratégie de mise en œuvre nationale exhaustive. Le Groupe Données devrait compter des représentants des milieux et organismes suivants :

- **le milieu de la recherche scientifique canadien;**
- **les organismes subventionnaires de la recherche canadiens (IRSC, CRSNG et CRSH);**
- **les fondations et fiducies de l'infrastructure de recherche du Canada (FCI, Génome Canada et les sociétés de fiducie provinciales équivalentes);**

- **les universités, collèges et instituts de recherche qui gèrent l'infrastructure de recherche du Canada;**
- **les ministères et laboratoires de recherche gouvernementaux (tant fédéraux que provinciaux) qui établissent la politique publique en matière de recherche;**
- **l'ABRC;**
- **l'Association des universités et collèges du Canada (AUCC);**
- **le CNC/CODATA;**
- **l'ICIST;**
- **BAC;**
- **Statistique Canada;**
- **CANARIE Inc.;**
- **le Commissaire à la protection de la vie privée du Canada;**
- **des représentants d'organisations d'étudiants et de professionnels qui participent à la planification de programmes de formation à la recherche;**
- **des représentants du grand public qui profiteront en fin de compte des bienfaits publics de la recherche financée par l'État.**

Le mandat du Groupe Données consisterait à orienter et superviser un petit secrétariat chargé de la mise en œuvre aux fins suivantes :

- **commander un projet pilote d'accès aux données (Projet Données) afin de faire la démonstration de concepts et des valeurs formulés dans ce rapport;**
- **planifier et superviser la formation d'une organisation permanente canadienne d'accès aux données (Données Canada);**
- **obtenir l'engagement à long terme de financement fédéral de Données Canada;**
- **élaborer un plan stratégique d'accès aux données (Plan d'accès aux données).**

8.2 Bâtir un soutien national

À l'échelon international, il existe un mouvement dont l'ampleur va croissant pour rendre les données de la recherche scientifique financée par des fonds publics le plus accessible possible afin de permettre la réalisation d'importants avantages économiques, écologiques, environnementaux, sociaux et sociétaux et accélérer considérablement la croissance des connaissances scientifiques. Cependant, la CNADRS a déterminé qu'il existe un manque de compréhension publique et de volonté politique, ce qui agit comme facteur inhibiteur à établissement de l'accès aux données scientifiques comme priorité nationale. Un soutien national est nécessaire pour mettre en œuvre les nombreux changements systémiques, juridiques, en matière de gestion, de régie et de changements financiers nécessaires pour que le Canada rejoigne les pays les plus dynamiques de l'OCDE en tant qu'acteur principal dans le nouveau contexte de recherche exigeant de grandes quantités de données. Il est manifeste que le Canada souhaitera faire partie des chefs de file et profiter des avantages qui leur permettent de se distinguer par un taux de croissance économique national supérieur, alimenté par l'innovation. L'absence actuelle de volonté se manifeste par le retard apporté à la mise en œuvre des recommandations conjointes émanant de la Consultation sur les archives nationales de données du CRSH et de BAC. Il n'a été possible d'établir des bases de données de recherche scientifique accessibles que dans le cadre d'initiatives nationales et internationales importantes mais relativement rares, notamment Génome Canada qui s'inscrit dans l'initiative mondiale sur le génome, la collectivité mondiale de recherche astronomique (DPOSS, SSGOT, etc.), le Centre for Global Research and Education in Environment and Health, la Bio-molecular Interaction Network

Database (BIND) et le module canadien (SCIB) du Système mondial d'information sur la biodiversité (SMIB). Il est important de bâtir sur de telles initiatives pionnières, de mobiliser les connaissances existantes dans toutes ces collectivités ainsi que dans les bibliothèques de recherche canadiennes, à l'ICIST et à Statistique Canada.

Recommandation 2 – Éducation

Que le Groupe Données, de concert avec les dirigeants des organismes parrains, commence immédiatement à sensibiliser les meneurs politiques et institutionnels et les guides d'opinion publique aux facteurs suivants :

- **le changement de contexte provoqué par l'accès à des ressources de données massives qui se produit dans le secteur de la recherche scientifique, dans le monde entier et au Canada;**
- **la nécessité pour le Canada de figurer parmi les chefs de file mondiaux de la transformation de la recherche, afin de conserver et de renforcer notre compétitivité économique et notre excellence scientifique à long terme;**
- **les avantages sociaux, médicaux, écologiques, environnementaux et économiques qui découleront de l'accélération du rythme de la découverte scientifique;**
- **les avantages éducatifs découlant de la capacité de situer l'apprentissage dans un contexte réel et de l'accès libre aux données;**
- **la nécessité d'une action concertée afin d'alimenter les changements culturels, juridiques, politiques et de gestion essentiels à l'établissement de Données Canada.**

8.3 Projet pilote

La mise en œuvre d'un système national pouvant appuyer la saisie, le stockage, le classement, l'accès et l'archivage de toutes les bases et de tous les ensembles de données de recherche dans le cadre d'un processus de gestion intégré et d'une structure de régie englobant de nombreuses institutions et organisations diverses est manifestement une tâche trop importante pour qu'elle puisse être accomplie sur-le-champ. Faute de preuves que le Canada a la capacité de mettre en œuvre un projet comportant tous les éléments nécessaires et de montrer les avantages et l'optimisation qui en découlent, une mise en œuvre intégrale ne sera pas possible. Ceci est particulièrement vrai si l'on tient compte des coûts éventuels, des besoins de ressources, des liens organisationnels, des avantages, des normes nationales et mondiales, des cadres de collaboration juridiques et internationaux ainsi que des solutions technologiques et de leur évolution, qui tous doivent encore être examinés et clarifiés. Cette complexité de la « vision d'ensemble » exige une approche plus évolutionnaire qui permet d'apprendre en progressant, ce qu'a recommandée la CNADRS.

Par ailleurs, la mise en place de Données Canada exige une saine coordination. Il existe, en plus des responsabilités stratégiques et de la coordination d'ensemble, qui seront au départ assumés par le Groupe Données et plus tard transférés à Données Canada, des secteurs pour lesquels une expertise particulière est nécessaire pour accepter, approuver et coordonner les activités. Ces secteurs sont notamment :

- l'élaboration d'une politique nationale sur les données scientifiques;
- des solutions technologiques pour une infrastructure évolutive d'environnements de traitement et de stockage de données, de systèmes d'accès et de gestion, de pratiques et protocoles hétérogènes;

- des protocoles administratifs pour la gestion de données pendant la totalité de leur durée utile;
- l'évolution des normes et pratiques en matière de métadonnées;
- des normes de qualité et d'intégrité des données et les processus et procédures concomitants d'évaluation.

Recommandation 3 – Financement

Qu'un financement suffisant soit fourni au Groupe Données afin de l'aider à mettre en œuvre un projet pilote de base de données à accès libre en 2005. Ce projet devrait être conçu de façon à montrer de quelle façon des bases de données, lorsqu'elles sont reliées, peuvent amener d'importantes percées sur le plan des connaissances. Il comporterait également des solutions aux difficultés posées par des facteurs tels que les politiques, la technologie, l'infrastructure, la gestion des systèmes, la qualité des données et des métadonnées, l'intégrité et la sécurité. Ceci ne signifie pas qu'il faut lancer un nouveau projet de recherche scientifique, mais il s'agit plutôt de compiler et d'intégrer un exemple convaincant qui est déjà en cours ou terminé.

Cette recommandation doit montrer à la fois la capacité du Canada à établir l'infrastructure et les processus de gestion intégrés nécessaires et la puissance de l'intégration d'une base de données interdisciplinaire en tant que méthodologie davantage efficace pour la recherche scientifique. Il serait fortement avantageux si le projet choisi complétait les activités de recherche actuelles exigeant de grandes quantités de données. Bon nombre ont déjà créé des bases de données relativement ouvertes et accessibles, et le projet pilote aiderait quelques-uns de ces investissements antérieurs. Le Groupe de gestion de projet (GGP) de la CNADRS propose de choisir un projet ciblant des données environnementales afin de concentrer l'attention sur les défis en matière de gestion des données dans le secteur de la durabilité environnementale. Un tel projet pourrait impliquer tous les conseils subventionnaires, la FCI et le CNRC, complèterait le SMIB et prendrait appui sur son expérience internationale. Il pourrait également encourager la modification de ce projet existant et fournir la preuve du leadership canadien.

Le GGP est également persuadé qu'un budget de l'ordre de 200 000 $, fourni par les trois conseils subventionnaires, la FCI, le CNRC et BAC, permettrait d'établir un tel projet de démonstration avec un financement d'un an. Il recommande également le financement pour la première année à hauteur de 100 000 $ d'un petit secrétariat chargé de diriger ce projet, afin d'aider le Groupe Données à s'acquitter des trois obligations mentionnées ci-dessus età établir le fondement pour la gestion opérationnelle de Données Canada.

Responsabilités de Données Canada

8.4 Capacité de gestion

Les participants de la CNADRS ont déterminé que le Canada devait être proactif dans le milieu international qui s'intéresse de plus en plus au libre accès aux données scientifiques et à leur conservation. Comme le Canada produit à peine 4,1 %[2] des connaissances scientifiques mondiales – selon le pourcentage annuel de documents de recherche produits – et représente 5,6 %[3] du 1 %

supérieur des documents classés en fonction des citations, notre capacité à tirer profit de découvertes scientifiques obtenues en partenariats, fondées sur des données mondiales, est plus importante que celle de la majorité de nos grands concurrents internationaux. L'accroissement des contributions canadiennes aux connaissances ne peuvent être réalisées qu'en ayant accès à toutes les données mondiales pertinentes disponibles et en les utilisant à l'appui du règlement des problèmes complexes que nous devons à présent étudier. Ce changement de pratique augmentera notre dépendance à l'établissement de normes mondiales. Par ailleurs, l'histoire montre que le Canada peut contribuer à influencer l'adoption de normes, car nous sommes principalement considérés comme un arbitre compétent et bien informé sur le plan technique d'approches, de normes et de protocoles concurrentiels. Le rapport du Priority Area Assessment on Data and Information du CIUS, intitulé *Scientific Data and Information*, argumente énergiquement en faveur de la création de mesures internationales relativement à un cadre stratégique à long terme pour les données et l'information scientifiques. Il recommande la formation d'un forum sur les données et l'information scientifique (SciDIF) chargé de superviser la création de ce cadre. Le Canada, par l'entremise de Données Canada et en collaboration avec le CNC/CODATA, devrait être représenté au sein du SciDIF.

Dans le cadre de la communauté mondiale du libre accès, il est essentiel que les milieux canadiens de la recherche participent à l'établissement d'accords internationaux sur des sujets particuliers, notamment le partage de protocoles, les normes concernant les métadonnées et le contrôle de bases et d'ensembles de données produits aux frais de l'État par des organismes publics ou privés ou des organisations non gouvernementales. Nous voulons également faire en sorte que les accords, normes et pratiques internationaux tiennent compte de la spécificité du Canada, notamment notre culture de la recherche, nos cadres juridiques, nos pratiques en matière d'accès aux données, l'information personnelle et la sécurité nationale.

Recommandation 4 – Participation internationale

Que Données Canada se dote d'une capacité de gestion qui lui permet de surveiller les forums internationaux sur l'accès libre et d'y intervenir afin de protéger les intérêts canadiens et aide la collectivité internationale à favoriser des accords, des normes et des politiques qui s'inscrivent à l'appui de pratiques exemplaires en matière d'accès, de partage et de sauvegarde, compatibles avec les besoins du Canada.

8.5 Protection de la vie privée

Certaines lois, particulièrement celles protégeant les renseignements personnels, notamment la *LPRPDE*, comportent de sérieuses implications pour l'accès à des données cruciales recueillies grâce à des deniers publics. Ces données ne doivent être utilisées qu'avec le consentement explicite, à une fin particulière et doivent habituellement être détruites lorsque leur fin a été réalisée. Cependant, il peut s'agir de données cruciales et, lorsqu'elles sont entièrement anonymisées, elles auront une valeur importante, particulièrement dans les domaines de la recherche en santé, démographique et sociale. À cause de la *LPRPDE* et de lois semblables, ainsi que de leur interprétation par les organismes d'application de la loi, de nombreuses sources de données sont perdues; des données qui pourraient avoir aidées à effectuer une recherche d'importance capitale ou à produire d'autres connaissances valables. Par conséquent, les participants de la CNADRS sont persuadés que des données entièrement anonymisées pourraient être sauvegardées, chaque fois que cela est possible, et rendus accessibles, en étant assujetties à des contrôles appropriés, pour de la

recherche pouvant être considérée comme une amélioration de la qualité de vie et de la santé des Canadiens, ainsi que de l'ensemble de la population mondiale.

Recommandation 5 – Éthique

Que Données Canada entame des consultations avec les commissions sur la protection de la vie privée, le Conseil national d'éthique en recherche chez l'humain, l'Association canadienne des comités d'éthique de la recherche, les bibliothécaires et archivistes de données et Statistique Canada afin de déterminer les obstacles juridiques à l'accès aux données scientifiques. De telles consultations devraient aboutir à des propositions de modification des lois sur la protection des renseignements personnels ou de leur interprétation juridique, afin de garantir que les données de grande valeur, financées par l'État et dont la confidentialité est convenablement protégée soient sauvegardées, sécurisées et rendues accessibles au moyen de mécanismes de contrôle de l'accès convenablement gérés. Par conséquent, les données pourraient être utilisées à des fins de recherche autres que celles qui était originalement prévues et pour lesquelles un consentement éclairé pourrait ne pas avoir été demandé.

Cette recommandation repose sur la prépondérance des intérêts du grand public sur les droits résiduels d'un quelconque particulier à l'égard de données anonymisées qui ont été recueillies en tenant compte des contraintes d'un consentement éclairé particulier. Cette recommandation ne s'applique qu'à l'information numérique et non à toute autre information ou « instanciation » physique.

8.6 Partage international de données

Des préoccupations juridiques et éthiques ont effectivement empêché la collaboration internationale et le partage des données issues de bases de données détenues par US functional Magnetic Resonance Imaging Data Center (fMRIDC). Le rapport du groupe de suivi de l'OCDE, intitulé *Issues of Access to Publicly Funded Research Data*, déclare que le fMRIDC a hésité à accepter des données d'endroits situés à l'extérieur des USA en raison des préoccupations relatives à la conformité aux normes de leurs Institutional Review Boards. Ce rapport prétend également que les chercheurs qui présentent ou demandent des données par delà les frontières nationales risquent de découvrir qu'il est particulièrement difficile d'agir en conformité avec les diverses lignes directrices éthiques qui existent dans différents pays. Il est évident que la pression pour une harmonisation des lois sur la protection des renseignements personnels entre les pays ayant des conceptions semblables de la vie privée augmentera, et qu'alors, il sera possible de partager davantage de données critiques de santé et d'autres données recueillies en vertu de l'application des règles du consentement éclairé et d'y avoir accès par delà les frontières internationales, sans contrevenir aux règlement et lois conçues pour protéger les renseignements personnels. Le Canada aura sans nul doute réglé ses propres problèmes à l'égard du partage de données de santé et démographiques réputées être personnelles ou pouvant constituer une menace à la sécurité nationale s'il souhaite éviter d'entraver son taux de découverte scientifique.

Recommandation 6 – Protection des renseignements personnels

Données Canada devrait entreprendre un examen de la *Loi sur la protection des renseignements personnels et les documents électroniques* ainsi que d'autres textes législatifs connexes

afin de découvrir les incohérences qui empêcheraient le partage de données sur le plan international avec des pays dont les projets de recherche en collaboration et les pratiques de partage de bases de données devraient être élevés – les premières analyses devraient se concentrer sur les États-Unis et l'Union européenne. Données Canada devrait également collaborer avec les législateurs canadiens en matière de protection des renseignements personnels pour aligner cette loi de façon à permettre un partage de données entièrement conforme entre des pays particuliers.

8.7 Données d'importance nationale

Il existe des cas notoires dans lesquels des bases et des ensembles de données revêtant une importance nationale ont été perdus ou contaminés parce que le gardien des données ou l'archiviste (s'il y en avait un) avait omis de prendre des mesures suffisantes de conservation des données. Des pirates informatiques, des virus et des vers et d'autres attaques électroniques malveillantes peuvent également avoir causé cette perte. Pour renforcer la sécurité de tels biens nationaux, il est hautement désirable de placer une copie de sauvegarde de la base de données dans un système d'archives comportant une protection accrue.

Recommandation 7 – Archivage

Que les bases et les ensembles de données, définis par Données Canada (p. ex., au moyen d'un comité d'examen de pairs ou d'un comité spécial ou d'une société ou association professionnelle compétente) comme étant d'importance nationale, soient déposées et mises en sécurité à BAC.

8.8 Responsabilités et avantages

Les bases et ensembles de données clés compilés par les chercheurs ou des équipes de recherche peuvent faire l'objet d'un examen avant d'être certifiés comme source de données fiable. De tels examens sous-tendent certes la volonté de la communauté de chercheurs d'utiliser de telles données, mais leur réutilisation peut engendrer de la PI susceptible de bénéficier d'une protection légale et de nouvelles connaissances, ce qui pourrait aboutir à l'élaboration de produits, de biens ou de services. En vertu de la loi canadienne actuelle, une telle utilisation pourrait imposer une responsabilité au gardien des données et peut-être à un organisme subventionnaire, au gouvernement fournissant les ressources à un tel organisme et à l'employeur du ou des chercheurs. La collaboration internationale en constitution de bases et d'ensembles de données peut également créer une responsabilité pour les contributeurs étrangers aux termes de la loi canadienne et étendre la responsabilité canadienne aux systèmes juridiques d'autres pays.

Par ailleurs, les propriétaires de bases et d'ensembles de données compilés à un moment particulier ne peuvent pas prévoir la réutilisation de ces données, et l'incertitude fera naître des préoccupations au sujet d'une responsabilité future éventuelle d'origine imprévue. Il est donc crucial que la responsabilité des gardiens de données soit bien définie dans la loi. Dans la plus grande mesure possible, cette loi devrait également encourager un leadership canadien grâce à l'accès aux données scientifiques.

Recommandation 8 – Responsabilité

Que Données Canada constitue un groupe d'experts pour examiner les cadres juridiques canadien et internationaux relativement à la responsabilité pour les bases et les ensembles de données et lui demande de proposer un nouveau cadre juridique canadien compatible avec l'évolution des cadres juridiques internationaux. L'objet consisterait à équilibrer la responsabilité des gardiens de données et de leurs institutions avec les avantages sociaux résultant de l'accès libre à ces données, afin de protéger les gardiens contre la responsabilité découlant d'utilisations futures inattendues des données.

8.9 Accessibilité de données anonymisées

Comme il a été dit en 8.5 – Protection de la vie privée, les données engendrées à partir d'informations personnelles ou grâce à un accès particulier à d'autres bases de données confidentielles ou sécurisées doivent habituellement être détruites lorsqu'elles ont rempli leur but visé, autorisé au moyen d'un consentement éclairé ou d'une autre permission explicite. Cette destruction de données potentiellement valables représente un coût important pour la recherche, empêche d'autres études longitudiales d'impact et empêche l'accès à des données potentiellement valables au milieu de la recherche. L'information personnelle doit certes être soigneusement protégée, mais des données recueillies aux frais de l'État devraient être disponibles pour le plus grand bien possible du public. À l'avenir, les données qui ne peuvent plus être retracées jusqu'à une personne ou à partir desquelles on ne peut déduire l'identité d'un particulier ou d'un petit groupe, devraient devenir accessibles pour d'autres fins de recherche.

Recommandation 9 – Anonymisation

Qu'un groupe d'experts soit nommé et chargé d'examiner les problèmes juridiques afférents à l'anonymisation des données et aux pratiques en matière de données sécurisées qui empêcheraient l'invasion de la vie privée d'un particulier si elles étaient rendues accessibles pour d'autres recherches. Le groupe devrait établir des limites à l'applicabilité du consentement éclairé, lorsqu'aucune identification n'est possible ou que l'identité de particuliers ou de petits groupes ne peut être obtenue par déduction. Si des occasions d'autoriser la réutilisation de données anonymisées se présentaient, le groupe d'experts proposera des modifications appropriées aux pratiques juridiques et réglementaires.

8.10 Bases de données à risque

Les institutions et organismes gèrent d'énormes quantités de données qui croissent de façon exponentielle. Certaines bases de données dépassent déjà les centaines de téraoctets et certains dépôts institutionnels gèrent à présent des pétaoctets. Les données et les métadonnées connexes présentent l'éventail complet de qualité, intégrité, contrôle de la propriété, accessibilité, contrôle de la version et gestion de la conservation. Données Canada devrait se concentrer sciemment sur la création d'un service compétent d'accès aux données et de conservation pendant qu'il se développe, mais il est évident qu'il doit aussi recouvrer ou mettre à niveau des bases et des ensembles de données existants pour satisfaire aux exigences du libre accès et de la conservation. Il faut agir d'urgence dans le contexte de la recherche actuelle pour aborder la vulnérabilité et la mauvaise accessibilité des bases de données actives et non actives.

Recommandation 10 – Bases de données à risque

Que Données Canada établisse un fonds afin de sauvegarder et d'améliorer l'accessibilité de bases de données actuelles de grande valeur, « à risque » et/ou critiques désignées par des groupes d'examen de pairs comme présentant une importante valeur actuelle, future ou historique.

Pour comprendre l'importance de cette recommandation, il sera nécessaire de désigner les bases de données existantes ayant la plus forte valeur soit comme bien national ou comme source de données de base pour d'autres initiatives de recherche ainsi que d'évaluer jusqu'à quelle mesure elles sont « à risque ». Comme le taux de perte augmente rapidement, il est important de commencer au plus tôt de les identifier et de les classer par ordre de priorité, avant que d'autres données d'importance nationale ne soient perdues.

8.11 Normes

La conformité aux normes et les examens de qualité relèvent normalement du gardien des données et sont financés dans le cadre de l'élément « gestion de données » de la demande de subvention de recherche. Cependant, pour satisfaire l'ensemble de la collectivité, tout groupe d'utilisateurs peut également faire effectuer des examens en présentant une demande à Données Canada. Les coûts directs liés à des examens demandés de l'extérieur devraient être assumés par Données Canada. Par ailleurs, dans le cas de bases et d'ensembles de données évoluant dynamiquement, des examens de la conformité aux normes et de qualité peuvent être déclenchés chaque fois qu'une des conditions suivantes s'applique :

- la méthode ou le processus de saisie des données est changé;
- de nouveaux contributeurs sont ajoutés;
- des données ont été modifiées ou ajoutées au-delà d'un seuil défini.

Toute base de données qui ne respecte pas les normes de qualité et d'intégrité de l'examen par les pairs doit être désignée comme telle dans ses métadonnées, et tous les utilisateurs précédents connus doivent être avisés de cette réserve.

Recommandation 11 – Critères et qualité

Que Données Canada collabore avec ses partenaires de recherche pour établir une fonction au sein de Données Canada (et de ses équivalents internationaux) afin d'officialiser les critères d'évaluation de la qualité des données ainsi que de définir des processus permettant de mesurer la qualité et l'intégrité des données.

Responsabilités des organismes de financement

8.12 Formation des chercheurs

Des bases de données à accès libre imposent un certain nombre d'exigences supplémentaires aux chercheurs. De nouvelles responsabilités découleront de la structure architecturale des données, des responsabilités liées à la garde et au contrôle, à la gestion de l'accès, à la création de

métadonnées et à la maintenance, à l'examen par les pairs et à la conservation. Peu de chercheurs – principalement ceux qui sont déjà impliqués dans les principaux mégaprojets exigeant de fortes quantités de données – possèdent les connaissances nécessaires pour assumer ces responsabilités de conservation. Ce n'est en particulier que dans le cadre des projets de plus grande envergure qu'il a été possible d'impliquer des archivistes de données dès le départ, et peu de chercheurs ont donc pu apprendre de ceux possédant de si vastes connaissances.

Il est urgent de s'attaquer à cette carence en fournissant aux chercheurs, particulièrement aux chercheurs principaux, une formation au sujet de la gestion des données, des responsabilités en matière de propriétés, de normes concernant les métadonnées et de solutions de conservation.

Recommandation 12 – Formation des chercheurs

Que toutes les organisations qui financent la recherche scientifique fournissent un financement particulier pour la formation* de tous les chercheurs principaux‡ aux pratiques exemplaires en matière de sélection de bases de données, gestion, gestion des droits et conservation de données, normes de métadonnées et autres questions importantes, afin que l'accès et la sauvegarde puissent être intégrés aux plans d'acquisition et d'entreposage de données dès le départ.

* lorsque les organismes ne financent pas la formation, ils devraient collaborer avec leur ministère afin d'obtenir des financement et mettre en œuvre des programmes de formation pour tous les chercheurs principaux qui sont les bénéficiaires de subventions ou qui les demandent.

‡ y compris les autres chercheurs qui sont impliqués dans la création de bases et d'ensembles de données.

8.13 Plan de gestion des données

Un nombre restreint de grands projets en santé, biologie, géologie, science des matériaux et astronomie comprenaient la création de bases de données accessibles ciblées sur une discipline, mais seuls les demandeurs de subventions du CRSH sont tenus de présenter un plan de gestion des données. Même lorsque de pareils plans sont présentés au CRSH, les ressources humaines et financières pour mettre en œuvre ces plans ont été systématiquement refusées. Tous les conseils de recherche et organismes de financement doivent corriger cette lacune en imposant comme condition de la demande de subvention la présentation d'un plan de gestion des données et en exigeant que le comité d'examen et de décision inclue la qualité, l'intégrité, l'accessibilité et la rentabilité du plan de gestion des données comme l'un des critères de décision pour l'acceptation. Les plans de gestion des données élaborés afin de créer des bases de données nationales ou de les compléter ou comme contribution à des bases internationales devraient obtenir des notes plus élevées que ceux qui ne servent que les intérêts de projets de recherche indépendants. Le rendement par rapport au plan de gestion des données doit également devenir une exigence en matière de présentation de rapports et les décisions subséquentes en matière de subvention devraient tenir compte de la satisfaction préalable comme critère de décision.

Recommandation 13 – Plan de gestion des données

Que les conseils de recherche et tous les autres organismes de financement de la recherche du secteur public et les ministères exigent que les demandes de projet et de subvention comprennent un plan de gestion des données, et qu'ils désignent spécifiquement le financement qui garantira la qualité, l'intégrité, l'accessibilité et la reddition de compte. L'inclusion d'un plan bien élaboré pour l'acquisition, la gestion, l'accès et la préservation de données devrait être une condition du financement. L'adhésion à de tels plans devrait également devenir une mesure du rendement non concurrentielle pour les projets et le portail vers des demandes de subvention subséquentes. Les conseils devraient reconnaître ceux-ci comme des coûts supplémentaires aux activités principales des projets de recherche.

8.14 Financement de base

Les projets qui élaborent des bases et des ensembles de données dont la valeur dépasse leurs bassins immédiats et initiaux auront besoin de ressources leurs permettant de les rendre accessibles et conservables. Par ailleurs, les gardiens de bases de données polyvalentes auront besoin de ressources permanentes pour maintenir et mettre à niveau ces bases de données. Par conséquent, il est clair qu'un financement stable et non concurrentiel à long terme est nécessaire pour garantir que de telles bases de données bénéficient d'un soutien et demeurent accessibles sur de longues périodes de temps à partir de dépôts institutionnels ou de BAC.

Recommandation 14 – Ressources

Que les ministères et organismes gouvernementaux fédéraux et provinciaux qui financent la recherche scientifique établissent des affectations à long terme stables et non concurrentielles dans le budget central afin de fournir aux institutions, organismes et organisations de recherche les ressources dont elles ont besoin pour sauvegarder toutes les bases de données importantes (ayant une grande valeur historique, actuelle ou potentielle). Que le gouvernement fédéral fournisse également un financement supplémentaire et suffisant à BAC pour garantir la sauvegarde en archives à long terme de tous les bases et ensembles de données importants.

8.15 Examen par les pairs

L'assurance de la qualité et de l'intégrité est une difficulté permanente que l'on rencontre dans les bases de données. Lorsque la recherche scientifique utilise des données compilées uniquement pour un seul projet, l'aménagement des données est en règle générale conçu uniquement pour ce projet. La qualité des résultats dépend entièrement de la qualité des données, et les deux font partie du processus examen de validation par les pairs. Cependant, dans les cas où les données font partie d'une base publique de données, on assumera implicitement que les données sont de la meilleure qualité et libres d'inexactitudes systématiques ou aléatoires. S'il incombe évidemment aux chercheurs d'obtenir la garantie que les données conviennent et sont de qualité suffisante aux fins de la recherche, l'ensemble de la communauté des chercheurs, en plus du gardien des données, a pour tâche sous-jacente de garantir la qualité et l'intégrité des données.

Recommandation 15 – Examen par les pairs

Que les bases et ensembles de données en cours d'utilisation ou dont l'utilisation est prévue dans le cadre d'initiatives de recherche multiples, y compris leurs métadonnées, fassent l'objet d'un examen par les pairs et que l'évaluation fasse partie des métadonnées.

8.16 Délais

Les participants de la CNADRS ont reconnu que l'on créera des bases de données qui fourniront de nouvelles idées cruciales aux chercheurs. Ces derniers devraient pouvoir bénéficier d'un accès exclusif à de telles données pendant une durée raisonnable mais limitée, afin que les connaissances puissent être préparées pour publication avant que d'autres chercheurs entament leur propre processus de découverte. L'accès exclusif doit être limité dans le temps, et les données deviendront accessibles publiquement peu après. Cette accessibilité publique et finale devrait cependant être contrôlée et restreinte lorsque le partage contreviendrait à la loi canadienne ou à des protocoles et accords internationaux.

Recommandation 16 – Délais

Qu'en collaboration avec Données Canada, les organismes de financement et ministère établissent des limites à la durée pendant laquelle les gardiens des données peuvent refuser le libre accès à leurs bases de données. Ce délai devrait être juste et raisonnable dans les circonstances du moment. Après la période précisée, les bases de données doivent être rendues accessibles publiquement, sous réserve uniquement des contraintes imposées par la loi ou des protocoles et accords internationaux.

Responsabilités des universités et des chercheurs

8.17 Systèmes de récompense et de reconnaissance

Le rendement et les systèmes de récompense universitaires sont fondés sur l'excellence de l'enseignement, de la recherche et du rayonnement. Le rendement de la recherche est évalué en fonction de la qualité de la recherche, des antécédents en matière de publication et de citation, du nombre de brevets accordés, de l'obtention de subventions pour la recherche et de la conclusion de partenariats industriels. Dans le système actuel, il n'y a pas de reconnaissance pour le leadership en compilation de bases et d'ensembles de données à accès libre de grande valeur ou de contributions importantes à ceux-ci, ni pour l'élaboration d'outils qui améliorent la valeur des données (p. ex., la combinaison de bases de données ainsi que les capacités d'accès et d'exploration). Pour changer la culture de la recherche, le système de récompense universitaire doit comprendre la reconnaissance pour des contributions importantes aux données de la recherche scientifique et leur utilisation.

Recommandation 17 – Récompenses

Que les facultés des universités, le corps professoral et d'autres unités de recherche universitaires élargissent leur système de reconnaissance et de récompense à l'intention des chercheurs afin d'y inclure l'excellence des contributions aux données scientifiques et l'élaboration d'outils pour l'amélioration de la gestion et de l'utilisation des données en tant qu'indicateurs de rendement importants.

Pour mettre en évidence ce nouvel indicateur du rendement, les universités, les conseils subventionnaires, la FCI et d'autres organismes de financement de la recherche scientifique devraient envisager de collaborer avec des associations industrielles et professionnelles afin d'établir des prix d'excellence pour les contributions aux bases de données issues de la recherche scientifique. Des prix de ce genre démontrent au milieu de la recherche la valeur de telles activités et au grand public, la valeur que les lauréats ont créée grâce à l'élaboration de bases de données et à l'accélération des connaissances nouvelles qui en résulte.

8.18 Formation de spécialistes

Par suite du consensus qui s'est créé au sujet du besoin de spécialistes en conception, gestion, accès et conservation de bases de données scientifiques, la formation actuellement proposée aux experts en la matière ne suffit pas. Elle ne répondra pas à la demande qui découle de la mise en œuvre de bases de données à accès libre et qui augmente rapidement. Il est essentiel d'agir immédiatement pour fournir ces ressources expertes dans le meilleur délai.

Recommandation 18 – Création de spécialistes

Que les établissements d'enseignement postsecondaires accroissent leur inscription d'étudiants en sciences de l'information et l'enseignement de l'accès aux bases de données et de leur conservation afin de régler la pénurie de bibliothécaires, de gestionnaires, de conservateurs et d'archivistes numériques formés.

---

[1] « Scientific Data and Information ». Un rapport du Committee on Science Planning and Review du CIUS, daté de décembre 2004 – ISBN 0-930357-60-4, qui peut être téléchargé de l'adresse suivante : http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf

[2] UK office of Science and Technology (Department of Trade and Industry) Survey of National Research Productivity Metrics – PSA target metrics for the UK research base – Oct 2004 : see (http://www.ost.gov.uk/research/psa_target_metrics_oct2004.pdf); Section 2.03

[3] Ibid, Section 3.07

9 Carte routière vers l'accès

Phase	Mesure	Date cible	Responsabilité
0	La FCI, l'IRSC, BAC, le CNRC, le CRSNG, le CRSH conviennent de créer le Groupe Données	Avril 2005	CNADRS
1	Le Groupe Données est mis en place avec un budget	Juin 2005	CNADRS
1	Projet pilote de base de données à accès libre approuvé et financé par le Groupe Données	Octobre 2005	GD
2	Les principales bases de données historiques « à risque » sont désignées et classées par ordre de priorité	Décembre 2005	GD
2	Les universités annoncent une augmentation des objectifs d'inscription en sciences de l'information.	Février 2006	U
2	Achèvement des cours de formation/séminaires pilotes sur l'accès des données et la gestion de celles-ci à l'intention des chercheurs (planification terminée en novembre 2005)	Mars 2006	CS
2	Exigence relative au plan de gestion des données pour la demande de subvention en place	Mars 2006	CS
3	Données Canada est créé et doté d'un budget initial par Industrie Canada	Mars 2006	GD
3	Données Canada participe activement à des forums internationaux sur l'accès libre aux données	Mars 2006	DC
3	Le plan stratégique de Données Canada est approuvé par le gouvernement fédéral	Mars 2006	DC
3	Approbation des projets de recouvrement des bases de données « à risque » les plus prioritaires	Mars 2006	DC
3	Politique d'accès exclusif élaborée pour les propriétaires de données	Juillet 2006	CS/DC
3	Les systèmes de récompense des universités/facultés doivent comprendre la reconnaissance de l'établissement et de la gestion de bases de données comme contribution précieuse au rendement	Juillet 2006	U
3	Augmentation des inscriptions en sciences de l'information réalisée par les universités	Septembre 2006	U
3	Démarrage du programme national de formation des chercheurs à l'accès des données, à la gestion et à la sauvegarde de celles-ci	Septembre 2006	CS/DC
3	Lancement des consultations sur la protection des renseignements personnels	Septembre 2006	DC
3	Lancement de l'harmonisation de la protection des renseignements personnels au plan international	Octobre 2006	DC
3	Lancement de l'évaluation de la responsabilité des propriétaires de données	Octobre 2006	DC
3	Les projets pilotes de base de données à accès libre sont terminés et accessibles	Octobre 2006	DC
3	Le catalogue des bases de données constituant un bien national est terminé	Décembre 2006	DC
4	L'engagement de financer Données Canada est inclus dans le budget fédéral	Février 2007	IC
4	Recommandations au gouvernement fédéral en matière de responsabilité des propriétaires de données	Avril 2007	DC
4	Le programme national de formation des chercheurs à l'accès aux données, à leur gestion et à leur sauvegarde passe à l'étape durable	Septembre 2007	CS/DC
4	Les recommandations en matière de protection des renseignements personnels sont présentées au gouvernement fédéral et aux conseillers en matière de protection des renseignements personnels	Septembre 2007	DC
4	Recommandations en matière d'harmonisation de la protection des renseignements personnels au niveau international au gouvernement fédéral	Décembre 2007	DC
4	La duplication des bases de données constituant un bien national par BAC est terminée	Décembre 2007	DC/BAC

Nota : Abréviations du tableau

CNADRS – président de la CNADRS en collaboration avec les présidents des conseils subventionnaires, du FCI, du CNRC, de l'AUCC et avec l'aide du conseiller scientifique national auprès du premier ministre.

GC – Groupe Données CS – Conseils subventionnaires IC – Industrie Canada

U – Universités DC – Données Canada BAC – Bibliothèque et Archives Canada

10 Annexes

10.1 Cadre pour une vision pour 2010 – Amélioration de l'accès aux données de la recherche scientifique (novembre 2004)

Amélioration de l'accès aux données de la recherche scientifique
Cadre pour une vision pour 2010

Document d'information pour la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS)
Novembre 2004

Introduction

Ce document réunit de brefs articles ou descriptions des neuf réalisations ou occasions thématiques qui ont émané de la réunion du 11 juin 2004 du Groupe de travail supervisant la Consultation nationale sur l'accès aux données de la recherche scientifique (CNADRS). Pour bien comprendre ces descriptions, il faut les lire conjointement avec le rapport de la réunion qui se trouve sur le site Web de la CNADRS à l'adresse (http://ncasrd-cnadrs.scitech.gc.ca/home_e.shtml).

L'objet de la réunion du Groupe de travail consistait à examiner d'une façon préliminaire le processus devant être suivi durant le forum de la CNADRS prévu pour les 22 et 23 novembre 2004. L'un des objectifs particuliers de la réunion était de comprendre globalement quelles occasions seraient créer pour le Canada en garantissant la gérance des données de la recherche scientifique et l'accès à celles-ci. Pour ce faire, on a demandé aux membres du Groupe de travail d'imaginer l'année 2010 et les réalisations qu'il faudrait avoir accomplies d'ici là pour atteindre l'état souhaité d'accès aux données. Ces réalisations ou occasions ont été résumées, selon la façon dont elles étaient perçues, par les divers rédacteurs ou équipes désignés ci-dessous.

1	*Leadership international en gestion des données*	Peter Pennefather Paul Uhlir
2	*Élimination des frontières entre les disciplines*	John Spray
3	*Nouvelles méthodologies*	Ellsworth LeDrew
4	*Amélioration de l'éducation*	John ApSimon
5	*Gérance des données*	Chuck Humphrey
6	*Renforcement de la capacité pour l'archivage des données*	Chuck Humphrey
7	*Les effets du réseautage*	Guy Baillargeon
8	*Nouvelles découvertes*	Robyn Tamblyn (santé) Steven Jones (génétique/génomique) Andrew Pollard (génie/instruments) Ellsworth LeDrew (environnement)

1. Leadership international en gestion des données

Au-delà des pratiques exemplaires pour la gérance de l'accès aux données produites par la recherche scientifique financée par le gouvernement au Canada : Ouvrir la voie de l'accès mondial à la recherche publique.

Introduction

Vue d'ensemble

Si les connaissances produites par des systèmes de recherche nationaux sont diffusées largement, il s'agit d'un bien public mondial (Pang et al Bull, OMS 81 : 815-820, 2003). Le devoir des organismes de financement de la recherche scientifique soutenus publiquement va au-delà de la simple distribution systématique de fonds pour la recherche. Ils doivent être capables d'articuler une vision pour leurs programmes, de garantir que les priorités désignées progressent et de faire la démonstration explicite que les résultats de ces programmes servent le public. Ces rôles de gérance sont essentiels au maintien du soutien public.

Un enjeu crucial de la gérance consiste à garantir l'accès aux données et à l'information produites par la recherche financée par le gouvernement. Les obstacles technologiques à l'accès et au partage des données sur le plan mondial ne cessent de tomber, grâce à l'investissement national et multinational permanent dans la cyberinfrastructure consacrée à la production, à la gestion et à l'utilisation des données, et des cadres internationaux sont élaborés pour régler les préoccupations et les problèmes qui découlent de la promotion du libre accès aux données (Arzberger et al. Science 303 : 1777-78, 2003).

Nous examinons ici comment les conseils de recherche du Canada peuvent aider les chercheurs canadiens à assumer un leadership international en gérance des données que leurs programmes de subvention les aident à produire. Nous prétendons que cette gérance doit transcender la gérance des données non traitées en amont jusqu'à leur intégration en aval (données reçues et comprises) et à d'autres produits de la connaissance.

Nous affirmons que de nouvelles occasions se présentent pour la gestion des données de la recherche primaire; et que la découverte de nouvelles façons de gérer ces données à l'échelle d'un système de recherche est un secteur dans lequel les Canadiens peuvent exceller. Ceci pourrait aider le Canada à se doter de la capacité d'établir des normes d'excellence pour la gestion de l'accès aux données découlant de la recherche scientifique financée par l'État et de démontrer explicitement comment cette gestion peut améliorer le rendement de l'investissement public dans la recherche. Le soutien accordé à une stratégie nationale pour les nouveaux échanges de données scientifiques et l'élaboration en partenariat d'informations fondées sur ces données pourrait mener à l'avènement de biens et de services novateurs axés sur l'interprétation, l'intégration et la réutilisation des données dont la valeur est reconnue au plan national et international.

Contexte

Le Canada est une petite nation disposant de relativement peu de sources de financement de la recherche publique. Ces sources appuient cependant un vaste ensemble de projets de recherche qui engendrent un flot sans cesse croissant de données. Ces données sont continuellement transformées en information et en applications mondialement utiles. Cette habileté à transformer les données en information pourrait être appliquée plus généralement s'il existait des normes internationales pour rendre les données mondiales disponibles et utilisables globalement. Il est donc dans l'intérêt des systèmes de recherche scientifique canadiens, qui en ont la capacité, d'assurer le leadership en montrant comment les données de la recherche pourraient être rendues plus généralement disponibles.

Grâce à la Consultation nationale sur l'accès aux données de la recherche scientifique (http://ncasrd-cnadrs.scitech.gc.ca/ncasrdfall_f.shtm), nous avons à présent l'occasion de décrire un point de vue canadien de la façon dont les données de la recherche et l'information qu'elles produisent sont à l'heure actuelle gérées ici même et dans le monde entier. Nous avons également l'occasion d'examiner de nouvelles approches numériques (particulièrement celles impliquant l'Internet) de l'élaboration et de l'établissement de normes pour une infrastructure partagée permettant d'améliorer l'accès à ce patrimoine mondial et son utilisation.

Une telle infrastructure est nécessaire non seulement pour améliorer le rendement de l'investissement du Canada dans la recherche financée par l'État, mais également pour quantifier la production d'information de la recherche au Canada et, en fait, dans le monde entier. Ceci est nécessaire pour garantir que les priorités des systèmes de recherche progressent effectivement et pour déceler des écarts entre les besoins perçus par le gouvernement et le public et ce qui est effectivement produit.

Il existe également de nouvelles occasions qui exigent de nouvelles façons de faire les choses. Jusqu'à présent, une grande partie de la recherche financée par l'État au Canada a été le fait de chercheurs individuels (CP), auxquels nous pourrions faire appel pour expliquer ce qu'ils ont fait et où leurs données peuvent être trouvées. Cependant, de nouveaux modèles de recherche en collaboration, ciblés et programmatiques, sont rendus possibles grâce à des progrès rapides de la technologie de communication par l'Internet.

Parmi les exemples possibles, citons le mouvement d'exploitation libre de logiciels, les archives de données du domaine public et les réseaux fédérés de données à libre accès, les examens par les pairs communautaires ouverts, les sites Web de recherche en collaboration, les collaborations pour les expériences virtuelles, les observatoires virtuels et les revues à accès libre, entre autres. Envisagées ensemble, ces nouvelles capacités représentent des aspects d'une tendance plus vaste vers la production, à la fois officieuse et officielle, par les pairs, d'information dans un milieu réseauté ouvert et largement réparti. De telles activités sont fondées sur les principes que l'on pourrait caractériser plus précisément comme « fonds communs d'information scientifique publique », plutôt que comme information exclusive produite par des CP individuels, et qui témoigne de l'éthique collaborative qui imprègne une grande partie de la science financée publiquement. Ces développements ont donné naissance à des occasions sans précédent d'accélérer la recherche et de créer de la richesse en se fondant sur l'ajout de valeur aux données et sur l'information non exclusive produite grâce au financement par l'État (Uhlir, 2003).

Certains de ces efforts sont entièrement fournis par des bénévoles non rémunérés, mais les rôles de premier plan joués par les employés de HP et d'IBM dans l'élaboration de logiciels d'exploitation libre font partie de la stratégie de prospection de ces entreprises. La responsabilité pour la production et le stockage des données primaires découlant des projets de cyberscience est également répartie souvent entre de nombreuses organisations. Ceci présente de nouvelles difficultés et occasions pour décrire et diffuser les résultats de la recherche (Hey et Trefethen 2004, www.ecs.soton.ac.uk/~ajgh/DataDeluge(final).pdf).

Il devient évident que même pour la recherche effectuée par les CP, l'accès aux données de la recherche et l'information qui en découle ne signifie pas simplement de mettre l'interprétation des CP à la disposition des chercheurs sous la forme d'une publication ou d'entreposage des données brutes dans un classeur. Les données primaires proprement dites sont transformées, d'objets analogiques numérisés après production en objets de données qui sont numériques dès le départ. C'est une gageure de garantir que les supports sur lesquels les données numériques brutes sont stockées demeurent accessibles.

En outre, pour que l'on puisse avoir accès facilement aux données et pour que l'information soit transmise à tous ceux qui pourraient en profiter, il faut qu'il y ait une documentation appropriée sur les données disponibles, sur les raisons pour lesquelles elles ont été produites et la façon dont elles l'ont été, ainsi qu'une description des avantages qui pourraient découler de leur utilisation. Dans bien des cas, la gérance de cette documentation (métadonnées) est même moins développée que la gérance des données proprement dites, mais toute aussi importante. La nature de plus en plus numérique des données primaires fait naître de nouvelles occasions pour la collecte de métadonnées (www.ifla.org/II/metadata.htm)

Initiatives internationales en matière d'information sur la recherche dans le secteur public

Un examen des régimes d'accès aux données de chaque pays dépasse la portée de ce document d'information. Au niveau international cependant, les régimes d'accès aux données de la recherche financée grâce à des fonds publics n'ont,règle générale, pas été particulièrement ouverts. Il existe néanmoins plusieurs exemples notables de politique générale sur le libre accès, notamment, les « principes de Bromley » de 1991 sur l'échange intégral et libre de données de la recherche concernant le changement mondial aux États-unis, les « principes des Bermudes » de 1996 sur la divulgation de données sur la séquence du génome humain, les principes de diffusion des données scientifiques du CIUS-CODATA de 1997 et l'initiative ministérielle de l'OCDE de 2004 sur l'accès aux données de la recherche financée par des fonds publics.

Des exemples remarquables de libre accès ayant obtenu un succès particulier comprennent la banque génétique GenBank, l'Observatoire international virtuel (OIV) et le Système des systèmes globaux des observations de la Terre (SSGOT). Mais ceux-ci sont alimentés par des communautés de chercheurs hautement motivés et bien intégrés qui toutes considèrent que la mise en commun de données comporte un avantage net pour les programmes de recherches individuels et qui ont établis une infrastructure importante à l'appui de leurs objectifs communs.

Le mouvement de libre accès dans la publication de documents scientifiques stimule également un nouvel examen de la façon dont l'information scientifique engendrée par la recherche financée par

l'État est diffusée. Il s'agit certes d'un problème bien plus vaste que ce qui peut être examiné adéquatement dans ce document, mais il est utile de relever quelques-unes des plus récentes initiatives législatives et stratégiques qui donnent un aperçu de l'orientation des politiques à l'égard du partage des résultats de la recherche.

Ainsi, les NIH ont recommandé au Congrès un nouveau plan qui, s'il est adopté comme loi, imposerait que tous les documents produits au moyen de fonds des NIH soient mis gratuitement à la disposition des chercheurs par l'entremise de PubMed dans les 6 mois suivant leur publication. Le Royaume-Uni examine une autre solution suivant laquelle les bibliothèques institutionnelles seront chargées d'apporter leur concours à l'accès aux versions autoarchivées de documents publiés par des chercheurs dans ces institutions. Ces projets de loi exercent une pression sur les éditeurs afin qu'ils ajustent leurs modèles commerciaux et laissent plus de latitude aux auteurs pour diffuser leurs documents. (voir http://www.earlham.edu/~peters/fos/newsletter/08-02-04.htm).

Il existe à l'heure actuelle plus de 1 200 revues scientifiques qui sont disponibles sur Internet dans des conditions de libre accès, notamment certaines initiatives dignes de mention telles que la Public Library of Science et BioMed Central. Les principes stratégiques du libre accès aux revues financées par des deniers publics ont été publiés tant aux États-Unis qu'en Europe en 2003 dans les « principes de Bethesda » et la Déclaration de Berlin. En 2004, de nombreux éditeurs de revues de sociétés professionnelles ont produit les « principes de DC » (Farnham et Brinkeley Scientist, 18 (13) :8, 2004) qui reconnaissent également la nécessité d'un vaste accès à la documentation scientifique produite par la recherche financée par des fonds publics. Certaines initiatives semblables ont également été mises en place aux États-Unis pour les articles de revue avant publication et la cyberpublication (p. ex., Cornell arXiv, créé à l'origine pour la physique des hautes énergies et étendu à présent de façon à inclure d'autres secteurs de la physique, des mathématiques, de l'informatique et de la bio-informatique), pour des articles de recherche individuels et d'autres ressources d'information (p. ex., the Social Science Research Network, le MIT D-Space initiative) et pour les supports pédagogiques universitaires (p. ex., le OpenCourseWare du MIT).

Il existe une tendance consistant à cesser de simplement verser des conclusions et des interprétations dans les archives publiques au moyen de publications. Ces initiatives se concentrent sur la façon de rendre des documents accessibles au plus grand nombre grâce à des dépôts électroniques. Cela signifie qu'il existe une occasion de relier ces documents désormais librement accessibles à des dépôts contenant les données primaires utilisées pour produire les interprétations publiées de ces données.

Les initiatives de libre accès dans les pays industrialisés sont souvent conçues en tenant compte expressément des besoins des sociétés en transition. En outre, de nouvelles revues librement accessibles sont créées dans ces sociétés. Dans ce domaine, le Canada montre déjà la voie. Ainsi, le travail de John Willinsky à l'Université de Colombie-Britannique à Vancouver, qui a créé le Open Journal Systems pour diffuser une revue approuvée par des pairs est très prometteur pour étayer de tels efforts. Les projets Ptolemy et Bioline de l'Université de Toronto essayent d'accroître l'échange de documents et de données de recherche entre les pays riches en ressources et ceux qui en manquent.

Il existe également des initiatives qui méritent d'être mentionnées et dont l'objet est de relier l'interprétation de données et de documents, de façon à réduire les écarts entre la recherche et la théorie, la recherche et les politiques ainsi que la recherche et la pratique. Citons comme exemples

les initiatives Cochrane et Campbell qui se concentrent, respectivement, sur des documents de recherche médicale et pédagogique. Le Canada s'enorgueillit de posséder un certain nombre de centres de recherche de pointe sur le sujet, notamment le Global Centre for e-Health Innovation à Toronto, le Public Knowledge Project à Vancouver et le Centre de collaboration en recherche pour l'efficacité en diagnostic.

Un rôle de chef de file de la gérance mondiale de l'accès aux données de la recherche pour le Canada

Qu'est-ce que le Canada a à offrir?

Le Office of Science and Technology du Royaume-Uni a récemment publié une enquête intéressante (www.ost.gov.uk/research/psa_target_metrics.htm) sur les mesures de productivité de la recherche nationale. Elle montre que les chercheurs canadiens dépassent les normes de rendement des chercheurs dans d'autres pays du G-8. Lorsqu'ils sont normalisés au PIB, ces documents de recherche canadiens démontrent que les chercheurs canadiens se classent deuxième, juste après ceux du Royaume-Uni, pour ce qui est des citations. Les taux de citation et de publication par chercheur se classent également deuxième après ceux des chercheurs du Royaume-Uni, et en bien meilleure position que ceux des scientifiques américains. Le Canada est le premier des pays du G-8 pour ce qui est de l'impact de documents fréquemment cités.

Lorsqu'il est normalisé au PIB, le rendement de l'investissement du Canada en termes de publication et de citation est admirable. Sa production ne représente cependant qu'un faible pourcentage du produit mondial absolu de la recherche. Le Canada n'a produit que 4,6 % des trois quarts de million de documents de recherche produits annuellement dans le monde, et 5,8 % de la production du 1 % supérieur de documents classés en fonction des citations. Les Canadiens ne représentent que 0,5 % de la population mondiale et n'engendrent que 2,4 % du PIB mondial de 32 billions de dollars. Le Canada, dont le PIB se situe aux environs de 0,8 billion de dollars, investit 15 milliards de dollars (1,9 %) en dépenses intérieures brutes de R-D (DIRB). Cet investissement représente à peine 7 % des DIRB américaines.

Le Canada dépense la même fraction de son PIB pour la recherche dans les institutions publiques (0,8 %) que les É.-U., mais 2 % supplémentaires du PIB américain vont à la recherche privée, soit deux fois le pourcentage canadien et plus de 30 fois le montant en dollars (7,5 contre 220 milliards de dollars). Cet écart constitue un défi. Si les systèmes canadiens de recherche financés par l'État veulent faire en sorte que leurs investissements dans le développement de la capacité de recherche aboutissent à un leadership international et un avantage national, ils doivent encourager le développement dans des secteurs ciblés dans lesquels l'innovation peut compenser la taille relativement faible de cet investissement dans la recherche publique. Nous proposons que la gérance des données de la recherche et de l'information connexe soit l'un des secteurs pour lequel le Canada possède le potentiel d'exercer un leadership international et de réaliser un rendement considérable sur tout investissement public de ce genre.

Quels sont les intérêts canadiens?

Les Canadiens peuvent à juste titre être fiers de l'excellence de leurs chercheurs individuels. Par la même occasion cependant, il faut reconnaître que le Canada est un petit pays dont l'intérêt national distinct peut être submergé par celui de son voisin bien plus grand du Sud. En raison de la facilité d'accès du milieu scientifique américain, les chercheurs canadiens sont poussés à définir la réussite en termes américains. Ils font la concurrence à leurs collègues américains pour du « temps d'antenne » dans les mêmes revues qui exercent un impact important et qui sont basées surtout aux États-Unis, ainsi que pour résoudre les problématiques discutées lors de grandes conférences scientifiques qui se tiennent habituellement aux États-Unis. Si l'on mesure le rendement en terme de citations, il y aurait un préjugé favorable à l'égard des intérêts américains, puisqu'un fort pourcentage de documents proviennent de chercheurs qui travaillent aux États-Unis et que ceux-ci citeront, comme il est tout à fait raisonnable de le supposer, d'autres documents qui sont liés à leurs intérêts. Il incombe aux programmes canadiens de financement de la recherche de faire en sorte que l'effet d'entraînement qu'exerce la proximité du Canada des États-Unis ne compromette pas la nécessité de créer un système qui sert davantage les intérêts canadiens.

Malgré l'exceptionnelle productivité du Canada en matière de recherche, des tendances inquiétantes sont manifestes dans les statistiques de l'UK-OST. Le nombre de publications par chercheur et la part des publications mondiales diminuent. De nombreux indicateurs de la productivité et de la recherche comportent cependant un important écart entre les intrants et les extrants, ce qui peut signifier que ces tendances ne se sont pas encore stabilisées. De nombreux chercheurs sont acculés à l'inactivité par l'incapacité d'obtenir un financement pour d'excellentes demandes de subventions. De très bonnes demandes sont à l'heure actuelle rejetées presque systématiquement en raison d'un manque de fonds. Les programmes de recherche canadiens distribuent certes des bourses d'études et des bourses de recherche ainsi que des récompenses salariales fondées sur la capacité de publier des documents et de rédiger de très bonnes propositions de recherche, alors que des institutions canadiennes de recherche recrutent certes des professeurs en s'appuyant sur des critères semblables, mais le Canada est en train de perdre sa capacité à soutenir suffisamment ses précieux chercheurs.

Cet effondrement de notre « système de production » au niveau des capacités de recherche témoigne peut-être de notre tentative d'imiter les méthodes de recherche qui prévalent aux États-Unis. où l'ordre de grandeur des fonds disponibles est plusieurs fois plus important. Jusqu'à présent, les chercheurs établis ont réussi à faire davantage avec moins et de ne pas se laisser distancer. On constate cependant une tendance inquiétante au Canada qui est de soutenir de grands projets menés par des CP individuels afin de permettre à quelques chercheurs canadiens de « faire la concurrence » à un niveau comparable à celui de leurs collègues américains. Aux États-Unis, des sommes élevées sont versées aux personnes qui peuvent exercer un impact important (tel que défini par les organismes de financement et le milieu scientifique américain).

Pour pouvoir être appliquée au Canada, dont les ressources de financement sont limitées, cette approche assimilable à une « force écrasante » doit être hautement sélective. Cette approche sélective affaiblit la capacité de tirer pleinement parti du bassin établi de chercheurs canadiens qui occupent des postes financés publiquement (au niveau provincial), qui ont été créés dans l'espoir de pouvoir obtenir un plus grand soutien à la recherche des organismes subventionnaires fédéraux. En élaborant une infrastructure, définie en termes généraux, qui permette la production et l'élaboration d'informations scientifiques sures, mais réparties et recueillies en collaboration, les

organismes canadiens de financement de la recherche pourraient contribuer à s'acquitter de leurs responsabilités en matière de gérance de la capacité de recherche. Ceci permettrait l'élaboration de programmes de recherche davantage concertés qui pourraient tirer parti des ressources d'infrastructure de la recherche réparties au Canada.

Les nouveaux milieux de la recherche collaborative que l'on retrouve sur Internet peuvent constituer un secteur très important, dont le développement pourrait apporter des avantages considérables aux Canadiens et un rendement plus important sur les investissements publics en recherche. Une telle infrastructure permettrait d'appuyer à la fois la recherche individuelle et la recherche en collaboration au Canada, tout en fournissant de façon concomitante un milieu pour le recrutement de chercheurs du monde entier, afin de contribuer à faire avancer les secteurs de la recherche jugés importants pour les Canadiens. Si le Canada peut établir des normes d'excellence, les chercheurs pourront obtenir l'accès aux données qui sont produites dans le monde entier et s'attarder à la tâche rude, mais moins coûteuse, de les interpréter.

Une telle initiative exigera l'élaboration d'une stratégie nationale pour l'amélioration de l'accès en collaboration et rationalisée (accrédité) aux données et à l'information de la recherche. Il n'est pas nécessaire pour cela que les chercheurs individuels perdent le contrôle de l'accès à leurs propres données. Il s'agirait plutôt de simplement mettre en œuvre des normes et une infrastructure pour garantir que tous les résultats de la recherche financée par les deniers publics canadiens soient traités de façon appropriée, sans consommer plus de fonds publics qu'il n'est nécessaire, tout en permettant néanmoins une utilisation en collaboration ou répartie, si cela convient.

De quelle façon le Canada peut-il aider au développement de la recherche mondiale?

Une approche moins centrée sur les États-Unis pourrait, en plus de favoriser le développement des priorités de la recherche canadienne, assurer le leadership permettant d'aborder l'incroyable dichotomie entre les pays en termes de leurs contributions à la recherche mondiale. Alors que 10 % de la population mondiale contrôle 90 % du PIB, ces mêmes 10 % contrôlent 98 % de la production de la recherche mondiale (définie selon les publications approuvées par les pairs). L'importance de systèmes de recherche nationaux fiables pour le développement international, particulièrement en terme de chevauchement avec les systèmes nationaux de santé, de protection de l'environnement et de développement économique, est dorénavant largement reconnue et favorisée par des organismes internationaux tels l'OMS, le COHRD et l'OCDE, parmi d'autres.

Le Canada pourrait jouer un rôle de premier plan dans le développement scientifique mondial en favorisant l'innovation et la diffusion des stratégies de gérance des produits de la recherche qui fonctionnent n'importe où et de façon répartie et peuvent par conséquent lier l'expertise canadienne aux décideurs et experts locaux et internationaux. Le leadership canadien en matière d'élaboration de protocoles et de mécanismes novateurs acceptés internationalement pour gérer plus efficacement l'accès aux données et à l'information produites par la recherche financée par des fonds publics, combiné à l'expertise connexe des chercheurs canadiens, pourrait profiter d'un effort concerté des conseils de recherche canadiens, puisqu'il y aura des problèmes communs de gérance de l'information qui transcenderont les frontières des disciplines.

Comment promouvoir une méthode canadienne de gérance de l'information produite par les données de la recherche financée par des fonds publics?

Le modèle IKEA. qui a remporté un vif succès, pourrait fournir un exemple. Pour commencer, il s'agit de chercher dans le monde des solutions rentables à des besoins quotidiens. Cette analyse du produit est ensuite suivie par la traduction efficace de ces idées en solutions utiles de style IKEA. Les portails normalisés IKEA (magasins) sont accessibles dans le monde entier, et tous leurs produits y sont aménagés stratégiquement selon des plans d'étage normalisés et d'une façon simple qui démontre clairement leur utilité. Ceci encourage des visites à répétition et facilite l'appariement de solutions précises à des besoins particuliers. Ce concept du « portail » a été mis à l'essai sous forme de prototype avec des acheteurs suédois représentatifs, puis exporté à travers le monde.

Le Canada possède le potentiel de se tailler une place de leader international en gérance des données/connaissances de la recherche publique et en normalisation des façons de partager ces données/connaissances. La petite base de recherche fortement répartie du Canada, l'excellente qualité de ses chercheurs, sa population hautement éduquée et son expertise en infrastructure des télécommunications, la gestion des connaissances et l'extraction de ressources pourraient être optimisés afin d'élaborer une solution proprement canadienne au problème de la réglementation rationnelle de l'accès aux produits de la recherche financée par les fonds publics. Les solutions élaborées au Canada pourraient alors être exportées dans le monde d'une façon qui procure un avantage au secteur privé canadien en créant des liens avec d'importants nouveaux marchés naissants et en les desservant.

Une première étape : enquête sur les données de la recherche et les stratégies de gestion de l'information utilisées par les chercheurs canadiens

Avant qu'une « stratégie nationale pour la gérance de l'accès aux données de la recherche scientifique » puisse être élaborée, il faut effectuer une enquête exhaustive des pratiques exemplaires actuelles au Canada et dans le monde entier pour la gestion des données et de l'information produites par la recherche financée par des fonds publics. Comme les données scientifiques sont un élément essentiel de l'infrastructure de la recherche moderne, il convient d'adopter une perspective mondiale. Cette enquête et les mesures subséquentes qui pourront être prises exigeront une acceptation et une participation importante de la part des chercheurs individuels, si l'on veut parvenir à une compréhension complète et exacte. Pour cela, il faudra mettre au point des mesures incitatives appropriées.

Les chercheurs dépensent d'importantes ressources pour gérer le flot d'information qu'ils produisent. Grâce à la nouvelle orientation préconisée par diverses sources à l'égard des responsabilités des chercheurs qui devraient être en mesure de rendre compte de leurs données et de les archiver, le coût de cette activité pour le chercheur individuel augmentera sensiblement dans un proche avenir. Les « Principles and Recommendations for Sharing Publication-Related Data and Materials » (Science Editor, 26:92-93, 2003) produits par le « Committee on Responsibilities of Authorship in the Biological Sciences » de la US National Academy of Science aux É.-U. sont un exemple de lignes directrices qui pourraient bientôt devenir obligatoires. Ces principes et lignes

directrices sont raisonnables et soigneusement pensés. Cependant, leur mise en œuvre risque d'exiger un investissement considérable de la part des chercheurs en termes de formation, de temps et de ressources engagées.

Pour aider les chercheurs canadiens à s'acquitter de responsabilités de ce genre, les organismes nationaux de financement de la recherche pourraient fournir un stimulus et un incitatif puissant à l'implication dans une solution générale et gérée centralement de ce problème. Ainsi, l'enquête ne devrait pas seulement documenter les solutions particulières actuelles, mais examiner également quelles caractéristiques d'une stratégie générale de gérance des données de la recherche encourageraient une participation enthousiaste. L'enquête devrait également évaluer la sensibilisation générale des chercheurs à leurs responsabilités et définir des options pour rendre accessibles les données de la recherche.

L'enquête devrait tenir compte des nombreux types et formes de données qui sont caractérisées par des descripteurs de contenu et de contexte. En prenant comme point de départ un document de travail de Paul Wooters, nous avons élaboré une matrice (voir l'annexe A) qui pourrait guider l'élaboration d'un instrument d'enquête. Les sujets insérés dans les cellules de la matrice doivent représenter les aspects clés de cette cellule.

L'objet de cette enquête devrait consister à définir un ensemble de normes réalisables qui pourraient évoluer en un programme national de gérance des données. La participation à une telle enquête pourrait finalement devoir être établie comme une condition pour recevoir des fonds publics, mais il serait conseillé, au départ, de recruter tout d'abord un ensemble de bénévoles pour découvrir la meilleure façon de mener cette enquête et la façon la moins intrusive possible. La solution finale exigera une optimisation des régimes de gestion et d'élaboration de politiques, non seulement des données proprement dites, mais également des métadonnées décrivant les produits de la recherche.

Une suggestion qui est peut-être sujette à controverse sur la façon d'acquérir une telle information à moindre coût et avec précision, consisterait à utiliser une forme bénigne des logiciels d'espionnage qui infestent à l'heure actuelle les ordinateurs de nombreux chercheurs. Ce logiciel devrait être d'exploitation libre et sa fonction devrait être entièrement transparente. Parallèlement, des mesures visant à protéger la confidentialité devraient être intégrées, mais on nous garantit qu'un tel instrument serait facile à mettre au point. Des versions commerciales de tels programmes de gestion des métadonnées tels que NuGenesis (www.nugenesis.com) sont largement mis en œuvre à l'heure actuelle. Cependant, ces systèmes sont fondés sur des plateformes exclusives et coûteuses de logiciel.

Comme presque tous les chercheurs canadiens comptent sur le fait que les collections de données numériques, la représentation et les stratégies de stockage sont exécutées sur des ordinateurs accessibles par l'Internet, la solution directe consisterait à élaborer un produit informatique classique d'exploitation libre permettant la production transactionnelle de métadonnées et acheminant cette information dans une base de données sécurisée mais interrogeable. Cette base de données recueillerait l'information sur les données qui ont été produites, identifierait par qui et pour quelle raison, et indiquerait au moyen de quelle méthode et à quel l'endroit ces données sont stockées. Comme l'acquisition et le stockage de données sont un processus dynamique, il sera nécessaire de mettre trimestriellement à jour l'information pendant au moins 18 mois.

GÉRANCE de L'ACCÈS AUX DONNÉES/DE L'INFORMATION : GRILLE D'ACTIVITÉ (catégories modifiées à partir de Paul Wouters; www.kbn.gov.pl/GRV/abstracts/4-3.pdf)

		CONTEXTE				
CONTENU		**Technologique**	**Institutionnel/ de gestion**	**Financier et budgétaire**	**Juridique et stratégique**	**Culturel et comportemental**
Acteurs (participants à l'échange)	D'égal à égal	BD transactionnelle	Programme	Minimal Local	Troc	Collégial Propre à la discipline
	Archives/ entrepôts	BD transactionnelle Production de mémoire	Programme Institut national	Substantiel Local/national	Réglementé Organisé	Patrimoine Efficacité
	Échange coordonné	BD transactionnelle Production de mémoire Carte des transactions	Programme Institut national Réseau national	Substantiel National	Contractuel Motivé par les résultats	Objectifs communs Incitatifs
Agents	(Re)producteurs d'information (protocole, réactifs, programmes, ordinateurs)	Ordinateur virtuel Banque de réactifs	Normes IPO Achats en bloc	Substantiel Local/national	Intérêt mutuel Mandaté	Responsable Recherche Éthique professionnelle BLP
Support	Personne-affichage-personne (conférence et cartographie des connaissances	Visualisation interprétative	Éducatif Scientifique Évaluatif	Annonce Rendement de l'investissement	Contractuel Sommatif	Visions du monde Langues Base de connaissances
	Médiatisé (par TIC)	Exploration des données Logiciel Matériel	FSI Sécurité	Coûts de la large bande	Vie privée Propriété	Village planétaire
Type d'information	Vérification des hypothèses	Grand nombre d'ensembles de données particuliers	Axé sur les chercheurs			
	Données d'observation et les «omiques»	Nombre limité de grands ensembles de données	Axé sur le consortium			
	Officieux/amateur					
	Métadonnées	Liens vers des ensembles de données	Axé sur les normes			
Objet de l'information	Recherche/ découverte	Approches nouvelles			PI	
	Réglementation/ développement	Approches établies			Documentation	

2. Élimination des frontières entre les disciplines

La création de données de la recherche nationales d'archives facilitera les progrès suivants d'ici 2010 :

Recherche interdisciplinaire grâce au partage de données

Par exemple, les statisticiens et les professionnels de la santé travaillant ensemble sont mieux en mesure d'évaluer la fréquence et la répartition des maladies de la population en fonction de l'âge des malades, de leur emplacement (environnement) et du climat dans le temps (épidémiologie). Les maladies infectieuses peuvent être retracées en terme des mécanismes de prolifération, du taux et de la spécificité de la cible. Envisageons, pour donner un autre exemple, l'effet sur l'urbanisme. L'étude des changements de densité de la population dans différentes parties de Toronto au cours de la dernière décennie, combinées à la planification bien conçue de la ville, a abouti à la construction d'une nouvelle voie de chemin de fer vers un faubourg particulier avant que le système de transport routier soit saturé. Ceci a pu être réalisé parce que le bureau du recensement a partagé ces données avec les experts en informatique, les urbanistes, les ingénieurs et la société de chemin de fer.

Ressources des bases de données internationales comme partie d'un milieu informatique plus vaste

Une collaboration plus étroite entre les pays peut aider notre compréhension de la démographie. Ceci nous aide à comprendre les tendances de la migration (p. ex., du Tiers-Monde vers le monde occidental). L'introduction de maladies exotiques (p. ex., la grippe aviaire) par des populations/individus immigrants peut être retracée grâce aux ressources des bases de données internationales partagées (c.-à-d. la collaboration entre les organismes d'immigration/des douanes dans différents pays). Les bases de données internationales partagées peuvent également aider à assurer la sécurité nationale et internationale en nous donnant la capacité de pister des personnes et des groupes.

Accroissement de la collaboration

La disponibilité des données donnera naissance à de nouveaux projets de recherche conjoints entre disciplines (p. ex., la physique et la chimie, les mathématiques et le génie). De nombreux ensembles de données sont utiles à plus de personnes qu'uniquement le chercheur/utilisateur à la source qui a produit la base de données. Ainsi, une analyse de la répartition du plomb et du mercure dans le sol d'une région particulière par un géologue peut être utilisée par un épidémiologiste afin de pister certains symptômes chroniques chez l'homme. Faute d'une base commune de données, la profession médicale ne serait pas nécessairement en mesure d'avoir accès à cette information et de l'utiliser.

Milieux intégrés des sciences et des données

Les données scientifiques doivent être formatées de façon telle qu'un mode commun d'accès et de classification soit élaboré. D'ici 2010, ceci aura été fait au moyen de divers formats normalisés.

Un logiciel aura été conçu afin de reformater l'intrant en une série de modèles, dont un ensemble aura été créé afin de pouvoir accepter la plupart des formats de données utilisés par les différentes disciplines. Ceci aura été réalisé en constituant un groupe de travail spécial chargé de concevoir des formats mondialement acceptables et en demandant aux experts de la TI de mettre au point un nouveau logiciel permettant de convertir les différents formats en un nombre plus restreint de modèles. La collaboration internationale s'emploie à uniformiser la présentation des données et l'accès à celles-ci dans le monde entier.

Nouvelles études en métadonnées

L'intégration du générateur de données et du produit de données a été rationalisée grâce à de nouveaux protocoles de métadonnées découlant en grande partie des activités d'intégration décrites ci-dessus. Ceci a donné naissance à un nouveau secteur d'étude et d'informatique impliquant la classification et la description d'objets et de produits qui semblent ne pas avoir de points communs. Ceci a des implications pour le contrôle informatique de systèmes complexes (c.-à-d. la découverte de points communs entre des objets qui semblent hétérogènes).

Le chercheur citoyen

En 2010, le particulier aura accès à la plupart des bases de données au moyen de son ordinateur personnel. Les organismes de santé pourront par exemple inviter des personnes à participer à diverses études liées à l'exercice, au régime alimentaire, à l'environnement et au travail dans le temps (année, décennie). Les participants seront en mesure d'observer les résultats pendant qu'ils évoluent dans le temps. Ceci pourra influencer les choix de style de vie.

Actualité des données

Une question qui est soulevée par bon nombre des observations ci-dessus est de savoir si les données disponibles sont à jour. Pour de nombreuses opérations prévisionnelles, il est nécessaire que les tendances des données se manifestent sur une durée raisonnable (mois, année, décennie), selon le sujet, et jusqu'à maintenant. Pour certains systèmes (p. ex., la démographie) il vaut mieux disposer des données les plus récentes possibles. L'automatisation de la mise à jour de l'information devrait donc être encouragée (p. ex., les personnes qui changent de domicile sont automatiquement enregistrées comme un changement d'adresse/emplacement, etc. dans une base nationale de données informatique).

3. Nouvelles méthodes

Méthodes technologiques

Les méthodes de traitement de données évolueront de plus en plus vers l'exploration de données au moyen de bases de données liées géographiquement et thématiquement, réparties dans plusieurs centres de recherche d'un pays et entre différents pays, connectées par un réseau de données à grande vitesse. Les projets de pointe actuels comprennent la modélisation « au moment adéquat » dans laquelle l'extrant d'une itération sur une cellule est terminé à point nommé pour être entré dans une autre itération de cellule pour un processus différent, à un endroit physique différent. Cette idée s'articule autour d'un espace de données reposant sur un « proxel » ou données rastrées qui intègre les données observées, assimilées en temps réel, les processus adaptés à cet espace (notamment physique, social et psychologique) et des données simulées fondées sur l'évolution de ces processus (Ludwig et al., « Web-based modeling of energy, water and matter fluxes to support decision making in mesoscale catchment », Physics and Chemistry of the earth, 28 (2003) 621-634). Parmi les difficultés, citons le traitement des données et l'accès à celles-ci dans un langage utilisé dans toutes les disciplines pertinentes, l'assimilation et l'archivage des données à durée de vie critique, la connectivité massive entre des nœuds différents et l'inclusion d'intervenants dans le modelage physique et conceptuel complexe.

Parmi d'autres possibilités technologiques, mentionnons :

- des panoplies d'outils en direct pour l'extraction et l'analyse préliminaire de données qui seront transparentes pour l'utilisateur non technique, l'étudiant et les médias ;
- la synchronisation des instruments et processus informatiques à de nombreux nœuds ;
- la réplique de logiciels à de nombreux nœuds pour garantir la conformité de l'analyse ;
- le contrôle de la qualité des données observées assimilées en temps réel, des données simulées et des données non métriques.

Méthodes institutionnelles et de gestion

Les nouvelles méthodes doivent comprendre la transparence du langage de traitement (notamment le Langage de modélisation unifié) entre des disciplines différentes « pousser » des techniques de traitement des données, la coordination des protocoles d'échange de données, le codage pour le transfert de grandes archives d'information image et, ce qui est plus important encore, l'engagement financier envers un milieu technique et scientifique en évolution.

Méthodes financières et budgétaires

Les futures méthodes seront transnationales, telles les rapports entre l'Agence spatiale canadienne, l'ASE et la NASA. Les problèmes de la propriété des données et de la PI technologique se transforment en importants obstacles à l'accès liés aux conflits inhérents entre le soutien fourni au secteur privé par le gouvernement, par opposition aux services publics fournis par le gouvernement. Ainsi, les chercheurs travaillant avec les données canadiennes de RADARSAT, qui examinent le déplacement des glaces sur l'ensemble du bassin arctique, doivent assumer des dépenses d'acquisition qui grèvent leur budget, tandis que leurs collègues européens ont facilement accès à des données semblables.

Méthodes juridiques et stratégiques

Le gouvernement continuera d'être un fournisseur important d'information environnementale. Certains protocoles, notamment les accords avec l'Organisation météorologique mondiale, sont efficaces pour le partage de données, mais de nouveaux types de données, notamment l'imagerie hyperspectrale pour l'exploration géologique, sera fournie, au mieux, sur une base de recouvrement des coûts et, au pire, elle sera exclusive. Ceci pourrait constituer un obstacle important pour une analyse efficace des données environnementales.

Méthodes culturelles et comportementales

Des données de divers types et structures devront être intégrées à l'appui de la prise de décision. Cet éventail comprendra de l'imagerie satellitaire de haut niveau traitée avec une précision géocodée, ainsi qu'une évaluation subjective telle que des connaissances plus anciennes des changements des pratiques de pêche locales en réaction aux changements climatiques. Il faudra élaborer de nouveaux contextes de collaboration et communication efficaces afin d'inclure ces divers types de données à un processus d'analyse systématique. Les archives de données peuvent devoir comprendre des indices de « compétence » des fonds pour divers types d'applications.

Contenu et contexte canadiens

Les Canadiens ont apporté d'importantes contributions novatrices aux méthodes de gestion des données. Le Système canadien d'information pour l'environnement (SCIE http://www.cise-scie.ca/francais/notices.cfm) est un récent exemple digne de mention qui a transcendé plusieurs frontières traditionnelles. Le programme SHARCNET (http://www.sharcnet.ca/) offre de nouvelles occasions en archivage et analyse de données réparties. Ces efforts seront compromis par la politique canadienne de recouvrement intégral des coûts et de privatisation des données satellitaires environnementales. Nous avons perdu considérablement de bonne volonté dans le milieu international de la recherche.

4. Amélioration de l'enseignement

La création d'une approche canadienne commune de l'accès aux données de la recherche scientifique financée par l'État ouvrira d'énormes possibilités sur le plan de l'enseignement, particulièrement dans le secteur de l'éducation postsecondaire.

Les programmes traditionnels axés sur une discipline et impliquant l'utilisation de données scientifiques et de génie seront enrichis par les expériences en enseignement liées à l'approche transversale (intégrative) de la disponibilité des données de la recherche. Les étudiants auront facilement accès à des données qui jusqu'alors n'étaient que difficilement accessibles, et le potentiel d'élaboration de programmes nouveaux et novateurs sera largement répandu. Les superviseurs de la recherche seront en mesure de recommander une plus vaste gamme de problèmes de recherche grâce au nouvel outil.

Les universités (et les collèges) disposent de systèmes rigides pour l'inscription, l'élaboration de programmes et les profils d'études pour tous les programmes sanctionnés par des grades et des diplômes. L'introduction d'un nouvel ensemble de normes pour l'accès aux données exigera une nouvelle évaluation de certaines de ces pratiques. Les universités sont à l'heure actuelle réticentes à changer les processus, à moins qu'il soit clairement démontré qu'ils comportent des avantages manifestes pour les étudiants, les chercheurs et, bien entendu, tout ceci sans l'affectation de nouvelles ressources. À moins que d'importantes nouvelles ressources soient disponibles, il est probable que nous ne verrons aucun nouveau changement de programme, à moins qu'il n'ait aucune incidence sur les revenus.

Il incombe donc au Groupe de travail de faire la démonstration des avantages pour toutes les parties. Certains de ces avantages sont décrits ci-dessous, de même que certaines réflexions sur une stratégie d'implication.

L'étudiant

Manifestement, tout processus qui procure une expérience éducative plus vaste et plus efficace est souhaitable. L'accès facile aux données de la recherche aura une incidence directe sur le choix de projets et exposera les étudiants, dès le début de leur carrière en recherche, à la puissance des données accessibles en tant qu'outil de recherche. L'exposition aux données d'autres chercheurs renforcera ensuite le réseau de chercheurs à tous les niveaux et, nous l'espérons, entraînera une plus grande mobilité des étudiants, qu'elle soit réelle ou virtuelle.

Les unités devraient, en prenant en compte la disponibilité plus généralisée de données de la recherche utiles, tirer parti du potentiel de leurs stratégies pour le recrutement d'étudiants diplômés (dans les départements voulus), la nomination de jurys de thèse/comités consultatifs et pour rechercher énergiquement des occasions d'élaborer des programmes conjoints entre institutions sœurs et les laboratoires du gouvernement. Ainsi, la Ressource de bioinformatique Canada du CNRC est un excellent outil de recherche, riche en données, qui pourrait constituer la base de projets de recherche conjoints entre des universités et le CNRC.

Les étudiants obtiendraient facilement le soutien de leur institution pour désigner les mentors de recherche à partir d'un éventail d'autres institutions. Il faut évidemment faire des préparatifs pour

de nouvelles occasions. À l'heure actuelle, les étudiants qui entrent dans une école du cycle supérieur sont déjà confrontés à un ensemble redoutable d'exigences. Bien souvent, surtout en sciences et en génie, un programme de maîtrise lourdement chargé mène à un programme de doctorat. À un point quelconque le long de ce parcours, de nombreuses universités ont des séances d'orientation pour les nouveaux étudiants et ceux qui poursuivent leurs études. Je constate un besoin d'inclure les aspects « puissance des données accessibles » de leurs programmes à cette étape. Les procédures varient d'un établissement à l'autre. Les séances d'orientation peuvent être données en laboratoire, au département, à la faculté ou à l'université (ou pas du tout, ce que je soupçonne être souvent le cas). Il faudrait qu'un volet « accès aux données et archivage » à l'intention des étudiants diplômés fasse partie du processus d'orientation pour tous les étudiants diplômés, et qu'il soit assorti d'un programme de cours national préparé pour un tel cours (une demi-journée à une journée). Une version en ligne ou en personne pourrait être dirigée par l'ICIST et les partenaires institutionnels du Groupe de travail. Les conseils subventionnaires pourraient, dans le cadre de leur stratégie de Personnes hautement qualifiées, dont ils vantent largement le mérite, exiger des institutions qu'elles présentent une telle séance d'information à tous les nouveaux étudiants et, au besoin, aux étudiants diplômés en deuxième année. Les boursiers postdoctoraux pourraient également bénéficier d'une telle exposition, de même que les finissants du secondaire participant à des projets de recherche pendant leur dernière année. Un tel atelier serait, par nécessité, plutôt général, mais au moins les occasions seraient-elles exposées clairement, particulièrement si elles sont présentées dans la perspective de l'allégement d'une partie du fardeau de la recherche et de l'élargissement de la portée.

Les étudiants en congé peuvent apprendre de nouvelles compétences par osmose (aux pieds de leurs mentors) mais ceci peut être un exercice frustrant. L'existence d'une politique nationale bénéficiant du soutien entier de tous les établissements et organismes de recherche pourrait assurer le leadership pour alléger le fardeau en tirant parti de l'accès aux données à la recherche et pour comprendre la nécessité de l'archivage. Ainsi, outre les ateliers d'orientation institutionnels mentionnés ci-dessus, il est possible de fusionner d'autres activités précieuses, notamment les cours d'été (cf. les ateliers d'études avancées de l'OTAN); de brefs internats pour les étudiants au gouvernement et chez des partenaires industriels, tant au plan national qu'international. L'idée d'un cours de courte durée « Accès aux données pour les nuls », donné par un consortium d'universités ou les conseils subventionnaires ou des institutions sans relation de dépendance (p. ex., Vitesse, voir www.vitesse.ca <http://www.vitesse.ca>). Les gouvernements possèdent des quantités massives de données, dont une grande partie est censée être disponible publiquement. On pourrait envisager des internats et/ou des projets de recherche conjoints donnant accès à de telles données à des fins de recherche.

Le milieu de la recherche

La production d'une cohorte de diplômés possédant de la facilité et une compréhension de l'accès aux données archivées de la recherche aura un effet immédiat sur le milieu canadien de la recherche en produisant un groupe sensibilisé et compétent de chercheurs et d'enseignants universitaires de la prochaine génération dans tous les secteurs de la société, garantissant que les contribuables canadiens tirent un rendement encore meilleur de l'argent versé pour la recherche financée par l'État. Sur le plan politique, il s'agira également d'un outil puissant pour recueillir le maintien et l'extension du financement de la recherche dans nos institutions publiques.

Le plan « Réalisations 2010 » du Groupe de travail désigne de nombreux secteurs qui dépendront tous du fait que la prochaine génération de chercheurs dans des disciplines riches en données sera capable de comprendre complètement les ramifications et le potentiel d'une politique d'accès aux données. Les observations formulées dans le reste de ce document sur ces secteurs révèlent des ramifications délicates et complexes nécessaires pour des experts et, partant, les occasions d'enseignement.

Quelques pensées sur une stratégie d'implication liée aux occasions d'enseignement

Cette partie est présentée sous forme de points et pourrait chevaucher le document sur la stratégie de communication. J'établis ici une distinction entre la communication et l'implication. L'année 2010 n'est pas tellement éloignée si l'on doit d'ici là influencer de nombreux organismes publics et privés.

1. Informer les conseils subventionnaires et demander qu'ils s'impliquent. Renvoyer aux stratégies nouvelles et existantes de Personnes hautement qualifiées. Rechercher des parrains pour des « ateliers d'utilisation de données » (CRSNG, IRSC, CRSH, Génome Canada, FCSCA, FCI, programme RCE).
2. Informer les ministères fédéraux et provinciaux tributaires de la gestion des données. Demander un soutien pour les ateliers, les internats, les placements d'été ciblés, etc. (EC (SPE, SMC); SC; AAC; RNCan)
3. Stratégie universitaire.
   a. Informer les vice-recteurs à la recherche (organisations régionales)
   b. Informer les doyens des études avancées (organismes régionaux pour l'approbation de programmes et au niveau national, l'ACES)
   c. Interagir avec l'équivalent américain, le Council of Graduate Studies.
4. Informer le conseiller national des sciences des répercussions sur la production de Personnes hautement qualifiées.
5. Préparer un exposé normalisé pour les « champions ».

Conclusion

Un ensemble passionnant de possibilités se dégage pour produire des chercheurs canadiens sensibles et experts dans le secteur de la gestion des données de la recherche. Il est essentiel que ces possibilités soient examinées dans le meilleur délai si l'on souhaite répondre aux attentes pour 2010.

5. Gérance des données

La gérance des données consiste en l'application des normes et comportements des chercheurs qui sous-tendent les pratiques de partage et de conservation des données de la recherche. D'ici 2010, plusieurs développements se seront produits au Canada, correspondant à un déplacement culturel au sein de la collectivité des chercheurs qui accentuera la gérance des données et mettra en valeur le chercheur comme « bon citoyen ».

Ce déplacement culturel coïncide avec un changement fondamental de la formation de la prochaine génération de chercheurs. On aura enseigné aux étudiants, dans le cadre de la formation à l'éthique faisant partie des programmes de troisième cycle dans tout le Canada, l'importance et la valeur du partage des données en sciences. La difficulté que posent les prétentions de propriété des données à l'accès libre aux données de la recherche s'est transformée en une compréhension de la gérance des données et de son importance, tant pour l'accès aux données que pour leur préservation. Les étudiants auront appris que le partage des données est un principe fondamental de la science ouverte et qu'il permet la réplication, l'exploration, la découverte et la création de nouvelles données à partir d'anciennes.

Le partage des données est largement connu dans le milieu de la recherche comme étant l'équivalent du livre « ouvert » en comptabilité. Chacun peut examiner la preuve. Par ailleurs, le partage de données de la recherche financée par l'État est reconnu comme étant le choix éthique approprié, parce que ces données font partie du patrimoine numérique du pays et, en tant que tel, appartiennent à la société.

Les responsabilités pour l'entretien et la conservation à long terme des données font partie de l'initiative générale des Archives nationales de données visant à instaurer la conservation numérique dans les pratiques de gestion des données. Au moyen d'un modèle du cycle de vie de la gestion des données de la recherche, les chercheurs établissent, avec l'aide des Archives nationales de données, une chaîne hiérarchique pour l'entretien et la sauvegarde des données qu'ils gèrent pendant la durée de la recherche et la durée utile des données subséquemment à cette recherche.
Les universités incorporent l'importance de la création, de la gestion et de la sauvegarde de données à leurs procédures en matière de titularisation et de promotion. Fait partie de ce système de récompense universitaire la reconnaissance des contributions intellectuelles liées à la création et à la gestion des données. Les lignes directrices en matière de pratiques exemplaires en création, gestion et conservation de données sont disponibles pour aider les universités à évaluer le rendement des chercheurs et leurs contributions dans ce secteur. Parmi les facteurs utilisés pour récompenser les chercheurs figurent l'utilisation secondaire des données émanant de leur recherche, c'est-à-dire que le partage des données fait dorénavant partie intégrante de la structure de récompense de l'excellence en matière d'érudition.

Les chercheurs utilisent dorénavant les données existantes de façon plus imaginative et, par conséquent, produisent une plus grande valeur pour l'investissement original dans les données. Les disciplines récompensent les nouvelles réalisations imputables à l'utilisation novatrice de données existantes, ainsi que l'utilisation originale de données nouvelles et anciennes.

Les comités d'éthique de la recherche reçoivent couramment une formation de base en pratiques exemplaires de collecte et de gestion des données. Les membres de ces comités comprennent l'importance de la conservation, tant pour la protection de l'investissement public dans les données que

pour les implications en matière de protection de la vie privée lorsque des sujets humains sont impliqués dans ces données. Ces comités comprennent que la bonne conservation des données procure une plus grande protection aux sujets humains que la destruction pure et simple des données. Un bon archivage garantit que les sujets humains ne seront pas identifiés ou que leur identité ne sera pas divulguée, tout en faisant en sorte que les données soient disponibles pour d'autres évaluations scientifiques afin de protéger les sujets contre des méthodes de recherche de piètre qualité ou mal intentionnées.

Les normes en matière d'anonymisation des données ayant trait à des sujets humains ont été élaborées et des pratiques exemplaires ont été établies. Les Archives nationales de données donnent à l'intention des chercheurs des ateliers sur les approches et les méthodes d'anonymisation des données afin de favoriser les pratiques exemplaires et de garantir la protection de la vie privée.

6. Renforcement des capacités pour l'archivage de données

Les étudiants de toutes les disciplines sont encouragés à envisager l'archivage de données comme choix de carrière. Les Archives nationales de données ont besoin de connaissances spécialisées dans divers domaines d'étude et engagent du personnel pour combler ces postes spécialisés. Dans toutes les disciplines, on accorde des diplômes spécialisés qui comportent une base importante dans un domaine, accompagné de l'archivage de données au moyen de nouveaux programmes qui ont été créés en information bibliothécaire et étude d'archivage. Les étudiants ayant décroché un diplôme canadien dans ce domaine hybride sont recherchés pour leur formation et leur expertise par les pays du monde entier.

En sus des nouveaux programmes débouchant sur un diplôme qui aident à renforcer la capacité en archivage de données, les conseils subventionnaires du Canada ont offert aux étudiants des mesures incitatives pour suivre une formation dans ce domaine. Le soutien des trois conseils a été mobilisé grâce à une initiative de recherche financée conjointement en préservation des données de la recherche. Cette initiative apporte un soutien aux étudiants diplômés qui travaillent à l'obtention de leur diplôme tout en finançant la recherche destinée à engendrer de nouvelles connaissances sur l'archivage des données. Les échanges avec d'autres archives nationales de données reçoivent un soutien de façon à accroître les connaissances et les expériences de travail avec d'importantes sources de données internationales.

7. Effets de réseautage

L'effet de réseautage se produit lorsqu'un bien ou un service est plus précieux pour un utilisateur quand un plus grand nombre d'usagers adoptent le même bien ou service. Des exemples classiques sont le téléphone et le télécopieur. L'achat d'un télécopieur par une personne profite indirectement à d'autres qui possèdent déjà un télécopieur en les rendant tous plus utiles. Au-delà d'un certain nombre de personnes qui adoptent le bien ou le service au départ, qui est appelé la masse critique, l'effet de réseautage devient important et attire de plus en plus d'utilisateurs parce que la valeur pour chacun d'entre eux dépassera le coût qu'il doit payer. La valeur d'un réseau dépend du nombre d'utilisateurs du réseau (loi de Metcalfe). L'utilité de grands réseaux, particulièrement des réseaux sociaux, peut augmenter de façon exponentielle avec la taille du réseau (loi de Reed). Ensemble, ces deux lois expliquent en partie la popularité explosive du Web dans les années 90.

Voici quelques exemples d'effets de réseautage dans le contexte de l'objectif de l'année 2010 :

- le partage de volumes importants de données environnementales émanant d'études d'impact et de changements environnementaux interdisciplinaires;
- la capacité de contacter les chercheurs/experts auteurs des données pour avoir accès à leurs données et à leurs connaissances;
- sensibilisation accrue aux problèmes et méthodes de la recherche dans d'autres disciplines et l'incidence fortuite potentielle sur la « valeur du chercheur »;
- plus d'équipes de recherche interdisciplinaires;
- lancement de nouvelles entreprises fondées sur l'accès aux données publiques et leur exploitation commerciale;
- plus grande collaboration en découverte scientifique par suite de l'amélioration de l'accès aux données.

Sachant que les effets de réseautage pourraient avoir un effet synergétique et d'accélération sur l'adoption de nouvelles méthodes et pratiques, comment pouvons-nous en tirer parti pour surmonter les difficultés de mise en œuvre d'un système de collecte, de stockage, de traitement, de distribution et de conservation de données de la recherche scientifique avec une diligence raisonnable à l'égard des droits de propriété intellectuelle et de contrôles à l'accès appropriés?

Dans le contexte de l'objectif de l'an 2010 et des réalisations qui devraient avoir été réalisées d'ici là pour parvenir à l'état souhaité **d'accès amélioré aux données de la recherche scientifique**, comment les effets de réseautage pourraient-ils jouer un rôle? Pourraient-ils être provoqués ou déclenchés par des mesures ciblées?

Facteurs à prendre en ligne de compte dans l'effet de réseautage :

Technologiques

- Le Canada possède déjà une infrastructure technique robuste et souple qui doit être maintenue et agrandie davantage.
- Il reste beaucoup de travail à faire en matière d'interopérabilité des logiciels et protocoles et de l'adoption de normes pour les métadonnées et pour l'échange de données et la qualité des données. Ce travail doit être effectué au moyen d'une participation dynamique et concertée des Canadiens dans les forums internationaux.

Institutionnels et de gestion

- Le Canada ne dispose pas d'un organisme national chargé de sauvegarder et de cataloguer les données de la recherche scientifique et d'y fournir un accès systématique, efficace et pratique, ainsi que d'aider d'autres institutions à élaborer des politiques et méthodes propres à une discipline pour donner un accès transparent et libre à leurs données.
- Les institutions devraient mettre en place des politiques qui limitent la tendance à empêcher la circulation publique des données.

Financiers et budgétaires :

- Pour le simple principe de l'efficacité opérationnelle, les organismes subventionnaires et le gouvernement devraient s'employer à maximiser le rendement de l'investissement en favorisant la réutilisation de données et en fournissant la documentation voulue, les spécialistes et les installations de gestion des données efficaces. Le financement de la gestion à long terme des données devrait recevoir autant d'attention que le financement de la recherche proprement dite.

Juridiques et stratégiques :

- La restriction imposée à la réutilisation des données publiques par le milieu de la recherche doit être éliminée ou réduite au minimum.
- Les politiques canadiennes actuelles en matière de droit de propriété intellectuelle en interdisent fréquemment la réutilisation de données.
- Les pratiques de recouvrement des coûts des ministères et organismes fédéraux doivent être examinées à la lumière de ce qui se passe dans d'autres pays qui ont des politiques d'accès aux données moins restrictives.

Culturels et comportementaux :

- Améliorer la structure et les mécanismes de récompense afin de favoriser le libre accès et le partage par les chercheurs individuels. Si l'on rend plus attrayant de se joindre à un réseau et d'y contribuer, les individus s'apercevront des avantages que leur procure le partage. Le libre accès aux données primaires dans des entrepôts de données permanents aboutira également à des procédures de correction d'erreurs à long terme et durables.

8. Nouvelles découvertes

Génétique et génomique

Le milieu de la génomique a prouvé, de bien des façons, qu'il était un modèle pour l'accès aux données et le partage de données. Les entrepôts de données pour l'information sur l'ADN et les protéines, notamment Genbank (É.-U.) et l'EMBL (Europe) ont révolutionné la façon dont la recherche en biologie se fait de nos jours, et sans ces ressources, il est difficile d'imaginer comment le domaine naissant de la bio-informatique pourrait avoir vu le jour. Poussant plus loin cette école de pensée, Genbank a plus récemment constitué un entrepôt de traces qui permet de déposer les données brutes telles qu'obtenue des dispositifs de séquençage d'ADN. Cet entrepôt de traces continue d'attirer de nouvelles utilisations et des analyses scientifiques de ces données qui n'étaient certainement pas envisagées au moment de sa conception.

Il est également inévitable que les domaines de la génétique et de la génomique profiteront encore beaucoup d'une stratégie de dépôt et d'archivage continus de données. Une autre étape récente est la création d'archives pour entreposer les résultats expérimentaux d'expériences sur l'expression génétique et, ce qui est tout aussi important, une description exacte et normalisée de l'expérience proprement dite. La nature intrinsèque de ces données est telle que lorsqu'on a accès à de grands ensembles de données, on peut grouper et analyser de façon plus fiable les expériences subséquentes de chacun. Des subtilités peuvent se faire jour qui autrement n'auraient pu être décelées, par exemple, de faibles perturbations des niveaux des gènes en réaction à l'ajout d'agents pharmaceutiques qui étaient précédemment passés inaperçues peuvent mener à des idées sur les mécanismes d'une réaction indésirable particulière à un médicament. De même, l'information sur la variation génétique humaine, qui est souvent un produit secondaire d'autres études, peut facilement être déposée pour être utilisée par d'autres aux fins de la cartographie génétique, du dosage fonctionnel ou pour des études de bio-informatique. De telles analyses pourraient facilement être envisagées pour aider à déterminer rapidement des variations génétiques ou des mutations qui confèrent une vulnérabilité ou une résistance à des maladies ou infections génétiques.

Comme il est habituel dans l'étude de systèmes complexes tels que ceux que l'on trouve en biologie, de nombreuses expériences échouent, et les résultats ne sont pas publiés. La capacité de diffuser ces résultats négatifs peut être considérée comme un changement important de la façon dont les résultats scientifiques sont diffusés. L'impact économique d'être condamné à répéter de façon indépendante la même expérience parce qu'il n'y avait pas de façon évidente d'en communiquer les résultats au milieu de la recherche doit certainement être important et ne pas être limité non plus au domaine de la génétique et de la génomique.

Le milieu de la bio-informatique a également été un défenseur convaincu des logiciels d'exploitation libre. Cependant, le simple fait qu'un logiciel soit d'exploitation libre ne règle pas les problèmes relatifs à la façon d'avoir accès à une version particulière sur laquelle une analyse a été effectuée ainsi qu'à celui de la disponibilité continue du logiciel d'une source fiable et digne de confiance.

Comme de nombreux formats de données et de fichiers exigent un logiciel pour la lecture et l'analyse, il vaut également la peine de relever qu'une archive de logiciels sera tout aussi précieuse que les données qu'elle supporte.

Santé

Dans le milieu de la recherche en santé, il existe au moins trois grands groupes qui pourraient réaliser des progrès considérables et passionnants grâce à la disponibilité d'un entrepôt national de données de la recherche publique.

Il y a en premier lieu les chercheurs en santé de la population. Ils se consacrent à comprendre la dynamique de la santé des populations – les facteurs de risque et les issues des maladies. La recherche a permis de déterminer l'importance du tabagisme pour le cancer du poumon et les maladies cardiovasculaires, de l'exposition au DDT dans certains cas de cancer, du traitement au fluorure pour la prévention de la carie ainsi que les facteurs de risque des naissances prématurées et leurs conséquences négatives.

Il y ensuite les chercheurs en épidémiologie clinique. Il s'agit typiquement de cliniciens impliqués dans le traitement de malades présentant des problèmes de santé particuliers. Leurs travaux de recherche comportent habituellement des études qui établissent le pronostic de la maladie, les meilleures méthodes de diagnostic et de traitement de la maladie et l'efficacité de nouveaux et d'anciens traitements thérapeutiques. Les résultats améliorent directement les bienfaits pour leurs patients.

Enfin, il y a les chercheurs en politique de la santé et services de soins de santé. Ils étudient les effets des politiques et de la prestation du système de soins de santé sur l'accès aux soins, la qualité des soins et les résultats pour les malades; il s'agit de questions qui constituent de graves préoccupations pour les Canadiens et les gestionnaires de soins de santé. L'impact des partenariats privé – public sur la qualité et l'accessibilité des soins de santé, la participation aux coûts par les consommateurs pour l'utilisation de médicaments et de services médicaux, l'impact des listes d'attente sur la qualité des soins et des résultats et les programmes de dépistage et de vaccination publics du cancer du sein et de l'hépatite sont quelques-uns des principaux problèmes qu'ils abordent.

Malgré la diversité de la recherche, ces groupes de chercheurs ont un besoin en commun. Ils doivent réunir et suivre des groupes de personnes pendant une durée prolongée afin de déterminer les effets d'un agent, d'un test, d'une politique, d'une pratique de soins de santé particuliers sur les résultats pour la santé. Une partie de ce travail est expérimentale, mais la majeure partie de l'étude scientifique consiste en observation. Elle repose sur la collecte d'information personnelle et de santé dans le temps, habituellement à partir d'une variété de sources, et sur l'analyse des variations naturelles de la population pour aborder des questions de recherche pertinentes. Pour supporter des réponses efficaces et en temps utile aux nombreuses questions qui doivent être abordées, les chercheurs ont commencé à réunir des cohortes – des populations de personnes qui avaient une exposition commune (p. ex., qui travaillent dans des professions ayant de forts niveaux d'exposition à des agents potentiellement carcinogènes), des états de santé ou des maladies en commun (p. ex., grossesse, diabète, sclérose en plaque, accident vasculaire cérébral, cancer) ou des interventions communes (p. ex., greffes). À l'exception du cancer, pour lequel de nombreuses provinces établissent dans la loi une notification exhaustive de l'information, il n'existe aucune ligne directrice ou norme pour cette forme de regroupement de population aux fins de la recherche permanente dans d'autres groupes de population (habituellement désignée comme registre de maladies ou de population).

Il en est résulté une prolifération de registres locaux de maladies et de population ayant chacun sa propre infrastructure coûteuse. Peu d'entre eux sont suffisamment importants pour étudier les questions clés en temps voulu. La recherche sur des problèmes de santé moins communs n'est sûrement pas effectuée parce que l'effort nécessaire pour réunir la population prend beaucoup plus de temps qu'il n'est possible pour un seul groupe de recherche fonctionnant avec des enveloppes de financement de la recherche restreinte d'en disposer (p. ex., sclérose latérale amyotrophique). Tous les registres de patients et de population ont du mal à maintenir des données de recherche de grande qualité en raison des coûts permanents et de normes différentes exigées pour le consentement et la gestion des données.

Des archives nationales de données aideraient grandement à résoudre bon nombre de ces problèmes. Elles permettraient l'élaboration de registres nationaux de patients afin d'examiner les causes potentielles, les traitements, les pronostics et les résultats de problèmes de santé communs et peu communs, ainsi que de nouvelles procédures (p. ex., greffes de cellules des îlots pancréatiques). Elles faciliteraient la recherche de meilleure qualité en normalisant la saisie des données et les besoins de codage. Elles permettraient d'appliquer des solutions bien plus énergiques à l'accès aux données et à leur sécurité, particulièrement en santé, où la protection de la vie privée et la confidentialité sont des enjeux cruciaux. À l'heure actuelle, les groupes de chercheurs disposent rarement de ressources suffisantes pour mettre en œuvre l'infrastructure et des processus de sécurité des données de pointe.

Les chercheurs posséderaient, dans l'archivage normalisé durable de registres nationaux de la population, un moyen d'avoir accès à des populations suffisamment grandes pour qu'ils puissent élaborer de nouvelles méthodes permettant de déterminer des grappes de malades qui, par exemple, seraient plus susceptibles de réagir à un traitement particulier ou de subir des complications. Les chercheurs seront plus à même d'élaborer des règles de prévision qui leur permettront de mieux apparier le choix du traitement ou la stratégie de prévention pour une personne donnée, ce qui améliorera de façon spectaculaire la probabilité de réussite. Le moment est bien choisi pour agir. Les rapides progrès de la génomique présentent l'occasion d'examiner les déterminants génétiques et gènes-environnement complexes de l'expression de la maladie et de la réponse au traitement, ainsi qu'un modèle pour la gestion de données très sensibles (p. ex., la biobanque du Royaume-Uni). L'arrivée de dossiers électroniques de malades en soins de santé fournit le mécanisme informatique grâce auquel des règles de prévision complexes pour le choix de traitement peuvent être mises en œuvre rapidement, avec un minimum d'effort de la part des cliniciens, car les recommandations de traitement peuvent être produites par des calculs automatiques avec les données propres aux patients. L'initiative canadienne de dossiers médicaux électroniques interexploitables, dirigée par Canada Health Infoway Inc., créera des entrepôts cliniques répartis de données de santé normalisées afin d'améliorer la sécurité et la qualité des soins des Canadiens d'un océan à l'autre. Les occasions synergiques existent pour tirer parti de cet investissement en créant l'infrastructure parallèle qui permettrait aux chercheurs d'utiliser ces données afin de créer de nouvelles connaissances et de réintégrer de nouvelles connaissances dans le processus de prestation de soins. Des archives nationales de dossiers de santé longitudinaux dépersonnalisés pour les Canadiens permettront également d'élaborer de nouvelles connaissances et méthodes afin d'améliorer l'assurance de la qualité et la sécurité des patients, la surveillance de la santé publique, les enquêtes épidémiologiques et la planification en matière de santé. En fait, l'absence de données accessibles et de grande qualité sur les soins de santé a entravé le développement d'une communauté scientifique dynamique qui aurait pu engendrer les connaissances nécessaires pour aborder les nombreux problèmes de notre approche surannée datant du XXe siècle de la prestation des soins de santé. Il est temps d'agir.

Génie et instrumentation

Les facteurs déterminant de nombreux flux de données peuvent se trouver en génie et instrumentation (qui comprennent la physique, la chimie, toutes les branches du génie y compris la science des matériaux). L'évolution de la recherche interdisciplinaire, et ce que l'on peut attendre en 2010, est notamment déterminée par les facteurs suivants :

1. combinaison de la biologie et des microfluides (génie et physique) pour l'élaboration de « laboratoires sur des puces »;
2. la science des matériaux comprend dorénavant plus que la « métallurgie » : accès à des faisceaux à haute énergie, tels ceux que l'on trouve au Centre canadien de rayonnement synchrotron, permettra de comprendre de nouveaux matériaux et de nouvelles propriétés de matériaux;
3. selon Richard Feynman, « il y a plein de place au fond », et la révolution nanotechnologique doit son nouvel impact à la science multiphysicochimique, à l'informatique et à la fabrication;
4. les séismes, particulièrement ceux d'origine terrestre sur la côte ouest du Canada, sont une préoccupation constante. Les données recueillies à ce jour dépassent un téraoctet par an (http://www.seismo.nrcan.gc.ca/cndc/index_f.php) et la résolution plus fine des instruments ou la densité de la répartition exacerbera davantage les problèmes de stockage de données;
5. le projet North-East Pacific Time-series Undersea Networked Experiments sera le plus grand observatoire océanographique relié par câble du monde. NEPTUNE fournira aux chercheurs, aux éducateurs, aux décideurs et au grand public une nouvelle façon d'étudier et de comprendre les problèmes cruciaux à notre survie, notamment les séismes, la sauvegarde des poissons, le changement climatique et les sources d'énergie;
6. l'Institut Herzberg sur l'île de Vancouver maintient des bases de données pour de nombreuses expériences à grande échelle (http://cadcwww.hia.nrc.ca/), notamment les données devant être obtenues de NEPTUNE, ATLAS etc. et qui devraient dépasser 100 téraoctets par an;
7. l'Observatoire de Neutrino de Sudbury (ONS) recueille des données de ses 9 600 tubes photomultiplicateurs en permanence et produit environ 1 téraoctet de données brutes par an. Ces données sont uniques et doivent donc être sauvegardées pour les générations actuelles et futures de chercheurs afin de pouvoir être traitées de nouveau au fur et à mesure que la compréhension du détecteur s'améliore ou que de nouvelles théories seront élaborées.

De nouvelles découvertes ont besoin de nouvelles technologies, de nouveaux instruments (capteurs, actuateurs, algorithmes) qui soit reposent sur des principes scientifiques existants soit ont besoin de nouvelles découvertes en soi. Ainsi, SNO-LAB est mis en place afin d'élargir les expériences visant à répondre à des questions telles que la « matière noire » par exemple, qui exige la mise au point de nouveaux détecteurs ou capteurs (http://www.sno.phy.queensu.ca/public/queens/subatomic.html). D'autres exemples abondent et sont résumés de façon pratique à l'adresse http://www.ecs.soton.ac.uk/~ajgh/DataDeluge(final).pdf.

Lorsque l'on acquiert des données, au moyen de capteurs ou d'autres instruments, le nouveau traitement de ces données constitue un élément critique du progrès des sciences et de la technologie. Ainsi, les données stockées électroniquement deviennent la preuve des découvertes qui sont annoncées dans les revues scientifiques, les dossiers de santé etc. L'accès à l'annonce de découvertes faites à la lumière de nouvelles idées et leur réévaluation, particulièrement par rapport à des données qui sont uniques sont très importants. Ainsi, les aéronefs sont à présent vérifiés en vol,

transmettent des données cruciales sur les moteurs au fabricant au moyen, disons, de SITA (www.sita.aero) et de cette façon le bon fonctionnement du moteur peut être surveillé en permanence ou, en cas de panne, on peut remonter à l'événement qui a entraîné la panne (voir, par exemple, http://www.aviationnow.com/content/publication/om/200104/om77.htm ou http://www.scientificmonitoring.com/resources/papers/State %20of %20Play_ATEM %20EMS %20Oct03.pdf.) À l'avenir, ceci pourrait être élargi à la surveillance et au pronostic en direct qui permettra des interactions entre le sol et les ordinateurs embarqués afin de modifier automatiquement le vol pour le mener à bon port.

Environnement

Technologique
La modélisation environnementale a toujours été à la merci des cycles informatiques et de l'accès aux entrepôts de données. Les modèles de circulation générale (MCG) évoluent de façon à inclure un modèle océanique entièrement couplé et un paramétrage plus efficace des limites des terres émergées. À l'avenir, ils comprendront des rétroactions intégrales avec la biosphère. À un horizon plus lointain en prévoit l'inclusion de données non numériques, notamment des modèles économiques, des pratiques agricoles de précision et des indices psychologiques qui déterminent les changements de l'utilisation des sols et sont touchées par les tendances atmosphériques telles que l'augmentation des concentrations de gaz à effet de serre. Les modèles comprendront également les observations contemporaines grâce à une assimilation des données en temps réel. Des modèles régionaux qui s'approchent de ces « idéaux » fonctionnent avec des temps d'intégration qui sont près de la durée effective des processus, ce qui est incroyablement restrictif du point de vue de l'analyse à long terme. Il est manifeste que l'échelle d'amélioration des nouvelles technologies d'analyse et des technologies d'accès aux données doit être subite, non progressive.

Bon nombre de ces modèles complexes ont recours à des nœuds répartis d'analyse tant sur le plan de l'espace que de l'expertise. Par exemple, une expérience actuelle utilise des calculs à un certain moment x à un endroit par un physicien spécialiste de l'atmosphère et exige des calculs à un moment x + y à un autre endroit physique par un phytophysiologiste dans un institut différent, et ainsi de suite, avant que le physicien spécialiste de l'atmosphère puisse continuer avec les itérations suivantes. Les problèmes de matériel, de logiciel et de contrôle de la qualité des données d'un tel arrangement sont importants.

Institutionnel et de gestion
Les simulations qui font appel à un apport d'instituts séparés par des frontières politiques sont déjà une réalité. Les problèmes de propriété des données, de propriété intellectuelle et d'adhésion aux normes doivent encore être abordés, mais pourraient empêcher l'exécution de certain types d'expériences qui devraient être effectuées.

Culturel et comportemental
Il existe une longue tradition de collaboration dans le milieu des sciences de l'atmosphère, mais nous avons toujours le modèle NCAR, le modèle Hadley Centre, etc. qui sont plus fondés sur des traditions sociales et le contrôle gouvernemental que sur un investissement ciblé, effectif et rationel de fonds et d'expertise.

Contenu et contexte canadien
Malgré ce qui précède, le Canada présente des antécédents internationaux enviables en modélisation et surveillance environnementales. Le groupe de modélisation du climat à l'Université de Victoria a apporté des contributions de dimension internationale à la modélisation couplée atmosphère, océan, glace et terre. Nous avons une solide réputation en technologie et analyse de télédétection, car nous sommes un chef de file mondial en technologie des stations terrestres, technologie de radar satellitaire et de la nouvelle technologie d'imagerie hyperspectrale, pour les plateformes sur satellite et aéronef. Nous possédons l'un des plus grands superordinateurs pour la modélisation. Nous n'avons pas encore fait de progrès innovateurs en gestion des données et en partage des données. L'attention exagérée qu'apporte le gouvernement à l'industrialisation et au recouvrement des coûts est une entrave importante.

10.2 OECD Final Report March 2003 (without appendices)

Promoting Access to Public Research Data for Scientific, Economic, and Social Development

OECD Follow Up Group on Issues of Access to Publicly Funded Research Data

Final Report

March 2003

Acknowledgements

Participants from the United States are being funded through NSF Grant ACI-9619020 to help coordinate this group's activities and develop a series of case studies to explore data access issues. Additional support for the group's activities comes from Netherlands' Ministry of Education, Culture and Science, and governments and agencies of the Follow-up Group's members.

The CODATA Secretariat hosted a meeting of the Follow-up Group in March 2002 in Paris. CODATA is the interdisciplinary Committee on Data for Science and Technology of the International Council of Scientific Unions (ICSU).

The Polish State Committee for Scientific Research hosted the 4th Global Research Village Conference (October 2002), which provided a forum to discuss the Follow-up Groups' Interim Report.

The European Science Foundation (ESF) is a partner in this activity. ESF hosted a meeting of representatives of member funding agencies in October 2002 in Strasbourg, France. The meeting provided a forum for discussion of the Interim Report of the Follow-up Group, and resulted in continued discussions at the funding agencies themselves and, in some cases, new policies (see Box 1). In addition, ESF helped solicit input on agency policies for this report.

Finally, The Ministry of Education, Culture and Science from The Netherlands funded a web survey of current data policies as well as a number of case studies. These efforts were directed by Paul Wouters of NIWI-KNAW, The Netherlands. Their publication in the series The Public Domain of Digital Research Data also contributed to the activities of this Follow-up Group on Issues of Access to Publicly Funded Research Data.

Executive Summary

It is now commonplace to say that information and communications technologies are rapidly transforming the world of research. We are only beginning to recognize, however, that management of the scientific enterprise must adapt if we, as a society, are to take full advantage of the knowledge and understanding generated by researchers. One of the most important areas of information and communication technology (ICT)-driven change is the emergence of e-science, briefly described as universal desktop access, via the Internet, to distributed resources, global collaboration, and the intellectual, analytical, and investigative output of the world's scientific community.

The vision of e-science is being realised in relation to the outputs of science, particularly journal articles and other forms of scholarly publication. This realisation extends less to research data, the raw material at the heart of the scientific process and the object of significant annual public investments.

Ensuring research data are easily accessible, so that they can be used as often and as widely as possible, is a matter of sound stewardship of public resources. Moreover, as research becomes increasingly global, there is a growing need to systematically address data access and sharing issues beyond national jurisdictions. The goals of this report and its recommendations are to ensure that both researchers and the public receive optimum returns on the public investments in research, and to build on the value chain of investments in research and research data.

To some extent, research data are shared today, often quite extensively within established networks, using both the latest technology and innovative management techniques. The Follow Up Group drew on the experiences of several of these networks to examine the roles and responsibilities of governments as they relate to data produced from publicly funded research. The objective was to seek good practices that can be used by national governments, international bodies, and scientists in other areas of research. In doing so, the Group developed an analytical framework for determining where further improvements can be made in the national and international organization, management, and regulation of research data.

The findings and recommendations presented here are based on the central principle that ***publicly funded research data should be openly available to the maximum extent possible***. Availability should be subject only to national security restrictions; protection of confidentiality and privacy; intellectual property rights; and time-limited exclusive use by principal investigators. Publicly funded research data are a public good, produced in the public interest. As such they should remain in the public realm. This does not preclude the subsequent commercialization of research results in patents and copyrights, or of the data themselves in databases, but it does mean that a copy of the data must be maintained and made openly accessible. Implicitly or explicitly, this principle is recognized by many of the world's leading scientific institutions, organizations, and agencies. Expanding the adoption of this principle to national and international stages will enable researchers, empower citizens and convey tremendous scientific, economic, and social benefits.

Evidence from the case studies and from other investigation undertaken for this report suggest that successful research data access and sharing arrangements, or regimes, share a number of key attributes and operating principles. These bring effective organization and management to the distribution and exchange of data. The key attributes include: openness; transparency of access and active dissemination; the assignment and assumption of formal responsibilities; interoperability; quality

control; operational efficiency and flexibility; respect for private intellectual property and other ethical and legal matters; accountability; and professionalism. Whether they are discipline-specific or issue oriented, national or international, the regimes that adhere to these operating principles reap the greatest returns from the use of research data.

There are five broad groups of issues that stand out in any examination of research data access and sharing regimes. The Follow Up Group used these as an analytical framework for examining the case studies that informed this report, and in doing so, came to several broad conclusions:

- Technological issues: Broad access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls;
- Institutional and managerial issues: While the core open access principle applies to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers;
- Financial and budgetary issues: Scientific data infrastructure requires continued, and dedicated, budgetary planning and appropriate financial support. The use of research data cannot be maximized if access, management, and preservation costs are an add-on or afterthought in research projects;
- Legal and policy issues: National laws and international agreements directly affect data access and sharing practices, despite the fact that they are often adopted without due consideration of the impact on the sharing of publicly funded research data;
- Cultural and behavioural issues: Appropriate reward structures are a necessary component for promoting data access and sharing practices. These apply to both those who produce and those who manage research data.

The case studies and other research conducted for this report suggest that concrete, beneficial actions can be taken by the different actors involved in making possible access to, and sharing of, publicly funded research data. This includes the OECD as an international organization with credibility and stature in the science policy area. The Follow Up Group recommends that the OECD consider the following:

- Put the issues of data access and sharing on the agenda of the next Ministerial meeting;
- In conjunction with relevant member country research organizations,
    - Conduct or coordinate a study to survey national laws and policies that affect data access and sharing practices;
    - Conduct or coordinate a study to compile model licensing agreements and templates for access to and sharing of publicly funded data;
- With the rapid advances in scientific communications made possible by recent developments in ICTs, there are many aspects of research data access and sharing that have not been addressed sufficiently by this report, would benefit from further study, and will need further clarification. Accordingly, further possible actions areas include:
    - Governments from OECD expand their policy frameworks of research data access and sharing to include data produced from a mixture of public and private funds;
    - OECD consider examinations of research data access and sharing to include issues of interacting with developing countries; and
    - OECD promote further research, including a comprehensive economic analysis of existing data access regimes, at both the national and research project or program levels.

National governments have a crucial role to play in promoting and supporting data accessibility since they provide the necessary resources, establish overall polices for data management, regulate matters such as the protection of confidentiality and privacy, and determine restrictions based on national security. Most importantly, national governments are responsible for major research support and funding organizations, and it is here that many of the managerial aspects of data sharing need to be addressed. Drawing on good practices worldwide, the Follow Up Group suggests that national governments should consider the following:

- Adopt and effectively implement the principle that data produced from publicly funded research should be openly available to the maximum extent possible;
- Encourage their research funding agencies and major data producing departments to work together to find ways to enhance access to statistical data, such as census materials and surveys;
- Adopt free access or marginal cost pricing policies for the dissemination of research-useful data produced by government departments and agencies;
- Analyze, assess, and monitor policies, programs, and management practices related to data access and sharing polices within their national research and research funding organizations.

The widespread national, international and cross-disciplinary sharing of research data is no longer a technological impossibility. Technology itself, however, will not fulfill the promise of e-science. Information and communication technologies provide the physical infrastructure. It is up to national governments, international agencies, research institutions, and scientists themselves to ensure that the institutional, financial and economic, legal, and cultural and behavioural aspects of data sharing are taken into account.

1. Preface

At its March 2001 meeting, the OECD Committee on Scientific and Technology Policy (CSTP) accepted a proposal from The Netherlands to establish a working group on issues of access to research information. The plans of the working group were presented at the October 2001 CSTP meeting. Subsequently, the Committee narrowed the scope of activities to access to and sharing of research data produced from public funding.[1] Participation in the group was broadened to include Australia, Canada, Denmark, Finland, Germany, Japan, Poland, the Netherlands, the United Kingdom, and the United States. The CSTP asked the working group to:

- Report on current practices concerning access to and sharing of research data and their underlying principles on the basis of case studies;
- Report on the effects of selected current data sharing practices on the quality of research and the progress of science;
- Suggest principles for making policy on data sharing within the relevant national and international policies and regulatory frameworks.

The report's core principle is that ***publicly funded research data should be openly available to the maximum extent possible***. Adoption of this principle will promote good stewardship of public knowledge, strong value chains of innovation, and maximize benefits from international cooperation (see Box 1). The report's findings and recommendations are addressed to: CSTP members as representatives from the governments of OECD member countries that carry responsibilities for national and international science policy and the functioning of research funding agencies; research institutes; and professional and scholarly associations. The objective is to contribute to a better understanding of the importance of research data access and sharing, and to offer suggestions on how the new digital challenges should be met.

Box 1: This core principle guides many public scientific institutions and scientists. However, it remains unevenly implemented. Most recently, it was adopted by the United Kingdom's Medical Research Council. After a workshop hosted by the European Science Foundation, the MRC drafted the following statement: MRC promotes the creation of a diverse range of datasets, many of which are rich in informational content, unique and cannot be readily replicated. Sharing allows scientists to extend the value of these datasets through new, high quality, ethical research and exploitation. It also reduces unnecessary duplication of data collection. Building preservation systematically into routine data management is part of good research practice: it strengthens quality, enables replication and audit, and provides a sound basis for data sharing. [2]

Building on a number of case studies and a great deal of other research, the report focuses on issues related to the access and sharing of publicly funded research data, in digital form, across all disciplines in the natural, health, and social sciences. Attention is paid to the international aspects of access and sharing relevant to scientific cooperation among OECD member states. Three significant topical areas fell outside the charge of this working group, however, and will require separate follow-up: issues particular to developing countries; issues related to data produced by a mixture of public and private funding; and the issue of national security restrictions in light of recent global events since 11 September 2001.[3]

2. Introduction

2.1. The changing information technology context for scientific research and innovation

Information and communication technologies (ICTs) are rapidly transforming research and the broader society: witness the growth in the number of Internet hosts per person, in the percentage of computers per household,[4] and in the continued rate of growth of chip, storage, and network technology capacity.[5] Concurrently, there has been an explosion in the amount of data produced across all types of scientific endeavour.[6] Continuing ICT advances, such as the development of grid computing, large-capacity optical transmission networks, wireless networks of sensors and devices, and complex imaging systems, promise to push these transformations farther and faster. ICT-dependent research, such as geographic information systems, data visualisation systems, and realistic modelling, are adding tremendously to our ability to study and understand the world in which we live. These developments provide researchers in OECD countries, and increasingly in developing countries, with the opportunity not only to be more efficient, more effective and better connected, but also to dramatically expand the scope and nature of their investigations.[7] Together they create the possibility of an "e-science infrastructure."[8] The growing activities in data collection, storage, processing, distribution, and preservation are, however, only loosely connected. They require systematic planning to realize the full potential of the emerging e-science infrastructure.

2.2. The benefits of data access and sharing in public research

Within this new technological context, more widespread and efficient access to and sharing of research data will have substantial benefits for public scientific research (see Box 2). Open access to, and sharing of, data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined.

Sharing and open access to publicly funded research data not only helps to maximize the research potential of new digital technologies and networks, but provides greater returns from the public investment in research.[10]

BOX 2: ACCESS to international data has helped produce a better understanding of public health issues and worldwide disease prevention and control. For instance, research on cholera outbreaks and their relationship to numerous environmental factors relied upon data drawn from epidemiology, NASA remote sensing, marine biology, microbiology, genomic data, and social science data. This research—an example of 'biocomplexity' studies supported by the U.S. National Science Foundation—would have been impossible without access to numerous databases. The effect of this interdisciplinary and international research project is an increased scientific and sociological understanding of cholera outbreaks and their prevention. [9]

Improving and expanding the open availability of public research data will help generate wealth through the downstream commercialisation of outputs, provide decision-makers with the necessary facts to address complex, often trans-national problems, and offer individuals the opportunity to better understand the social and physical world in which we all live (see Box 3).

As a key link in the value chain of investments in research, open access to factual data plays an increasingly important role in all these areas.

BOX 3: A recent analysis demonstrated the economic benefits of providing open access to government meteorological data without any restrictions on re-use. 11 The "value adding" meteorological information industry in the United States has revenues in excess of $500M annually. The public meteorological data also support a rapidly growing weather risk management industry that underwrites financial risk management instruments valued at approximately $8B. In contrast, the private-sector value adding industry for meteorological information in the European Union is very small, largely attributable to the highly restrictive data policies of most national governmental meteorological services. What are harder to measure, but certainly occur, are the countless lost opportunity costs for researchers, students, and various other potential public users who find the high costs of the public data to be too great to use.

2.3. Roles and responsibilities of governments

If researchers throughout the world are to take full advantage of ICTs to improve and expand access to, and sharing of, research data, existing technological, institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural aspects must be addressed comprehensively and in an integrated way. To date, these aspects have often been treated on an ad hoc, project-specific basis. Given that OECD countries spend tens of billions of dollars each year collecting data that can be used for research, and for other social and economic benefits, ensuring that these data are easily accessible so that they can be used as often and as widely as possible, is a matter of sound stewardship of public resources (see Box 4).

BOX 4: Poor stewardship and lost opportunity for data access is exemplified by the case of Statistics Canada, which attempted to recover costs for its data management by charging data users. The effect of this form of management of these public data was a dramatic decrease in their use. In a study of the case, it was found that "Cost recovery was supposed to introduce a market type discipline on the demand for and supply of goods and services provided by the government. Since in economic terms Statistics Canada's outputs are public goods, the type of discipline envisioned by this policy is impossible to attain. Instead we have users who complain, refuse to pay and generally attempt to find alternative sources for their information needs. This policy fails the improved management of resources test."[12]

Scientists, research institutions, and research funding agencies around the world are increasingly engaging in large-scale, data-intensive projects. Such projects require data-management infrastructure, data-exchange protocols and policy frameworks, and a broad professional understanding that more extensive availability and use of the data is both necessary and desirable. Over the past decade, numerous studies, disciplines, research programs, and agencies have begun to address the complexities and benefits of open data access and sharing arrangements.[13] As scientists become better connected with each other, particularly through the Internet, and as research focuses on issues of global importance, such as climate change, human health and biodiversity, there is growing need to systematically address data access and sharing issues beyond national jurisdictions and thereby create greater value from international co-operation. The goal should be to ensure that both researchers and the broader public receive the optimum return on public investments, and to build on the value chain of investments in research and research data.[14]

3. Core Principle and Premises

The findings and recommendations that follow are based on the central principle that:

Publicly funded research data should be openly available to the maximum extent possible

As a general principle, publicly funded research data should be as open as possible and available at the lowest possible access cost, subject only to legitimate restriction and considerations. Restrictions may be necessary for reasons of national security, for the protection of privacy of citizens, or the confidentiality of trade secrets. Access to research may be limited by the respect for private intellectual property rights. Finally, there may be reasons for granting temporary exclusive access to those who collected the data. But the guiding principle should be openness.

In order to derive the maximum benefit from public investments in research data, access, use, management and preservation must be an integral part of the research process. Conversely, data should not be considered an expendable by-product of research. In many cases, data have value beyond the project and anticipated use for which they were originally collected. The reuse of publicly funded data for research and other types of applications should be promoted and not restricted.

The accessing and sharing of data is not merely a technical matter, but also a complex social process in which researchers have to balance different pressures and interests. Purely regulatory approaches to data sharing are not likely to be successful without consideration of these factors. Various approaches to data access and sharing are therefore necessary, including the establishment of regulations and incentives, and the dissemination of best practices.[15]

The following three premises complement and support the core principle of this report:

3.1. Data from publicly funded research are a public good produced in the public interest
Both the data from publicly funded research and research itself have strong public good characteristics that support their open availability to the public, and especially to other researchers.[16]

3.2. Factual data are central to the scientific research process
The production, open dissemination, and unfettered use of factual data are essential attributes of, and inputs to, modern systems of scientific research and technological innovation. Recognizing the role of digital data as fundamental to the value chain of science, technology and innovation will enable an optimum return on public investments.

3.3. Data access and sharing issues are international in scope
To more fully exploit the possibilities of global digital networks, and to capture their benefits for the global community, policy issues concerning access to and sharing of publicly funded scientific research data must be addressed, not only at the institutional and national levels, but also at the international level.

4. Data Access Operating Principles and Attributes

Data access and sharing requires effective organization and management. The necessary components that make for this organization and management may be characterized as "data access regimes." In their ideal form, these regimes enable all participants in the scientific research process to freely and efficiently access and share data. Adequate data access regimes require dispersed, as well as centralised, responsibilities across different management domains that include the technological, institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural.

No single approach to developing an effective data access regime is possible; however, a list of operating principles for and attributes of effective data access regimes and resources can be offered. This list of attributes and operating principles is based on a broad set of experiences, and supported by the case studies conducted for this report. Key attributes are listed below, and illustrated with an example from the case studies. [17]

4.1. More explicit access regimes

There is a universal need for the formalisation of institutional rules and data management policies. This formalisation follows from the growing complexity and scale of scientific research and the increasing expenditure on research data. At the moment, it is not clear who is authorised to distribute data across the globe. To reach the necessary transparency in the tasks and responsibilities of those involved, terms of access and use of data that rest on tacit agreements will have to be made explicit and formalised. A systematic and institutionalized approach is needed to help address operating characteristics of data access, and to take advantage of the opportunities arising from publicly funded research.

4.2. Operating Principles

4.2.1 Openness.

Open availability of publicly funded research data to the maximum extent possible is the core principle of this report.

4.2.2. Transparency of access and active dissemination.

Open data access requires actively disseminating where the data can be found, what the context and structure of the data collection is (metadata), how long the resource will be accessible, and what protocols and standards are employed. In short, this principle refers to the systematic visibility and traceability of data resources.

4.2.3. Assignment and assumption of formal responsibility

Formal responsibility for tasks associated with data access must be assumed by the appropriate participants in the global science system. The various individuals and institutions involved in the chain of data-related activities all have specific manifest and latent duties and obligations. These are founded in formal legal and professional normative standards and in the regulations of various agencies. Responsibility must also be assumed for various rights in the data supply, such as authorship, producer credits, ownership, financial arrangements, licensing terms, and, where appropriate, restrictions on use.

4.2.4. Professionalism
Codes of conduct, and related normative standards, of professional scientists and their communities can help to promote good practice and simplify the regulatory aspect of access regimes.

4.2.5. Interoperability
Technical and software standards and protocols are required to ensure the access and usability of data. These should be clear to the user and adopted by as many data management organizations as possible.

4.2.6. Quality
Quality refers to the proper description of uncertainties surrounding the production of the data (e.g., the techniques employed in their collection and archiving, and the measuring instruments and their calibration), the ability to ensure that the cited source and value are authentic, that the data retain integrity (complete and absent from introduced errors), and that they are secure against loss, destruction, modification, and unauthorized access.

4.2.7 Operational Efficiency
Open access to data increases the efficiency of research by avoiding unnecessary duplication of data collection and permitting the creation of new data sets by combining data from multiple sources. Coupled with open access, comprehensive documentation of data sets and how to access them provides a more efficient use of resources.

4.2.8. Flexibility
In general, scientific communities will approach data management requirements more consistently within their discipline internationally, than they will across other disciplines on a national level. Data access regimes need to be sufficiently flexible to take account of this variation.

4.2.9. Property
Institutional intellectual property rights as well as the individual rights of researchers are considerations of property interests. Unlike the private sector, public research operates on a principle of collective property interests, which are promoted by the open access and sharing of data resources.

4.2.10. Legality
Legal restrictions may limit access to and use of data.[18] Restrictions will apply primarily to 'secondary' data sets compiled for purposes other than scientific research. In some cases, the sensitive parts of data sets can be left out without rendering them useless. Specific types of legal restrictions include: national security, privacy and the protection of trade secrets.

4.2.11. Accountability
Accountability involves measuring the cost, benefit, and performance of data access and sharing regimes and taking appropriate actions in response to the results.

4.3 Building a Data Access Regime: the Global Biodiversity Information Facility (GBIF)
The Global Biodiversity Information Facility (GBIF), which began under the auspices of the OECD Megascience Forum, has sought to implement these principles as a means to achieve the larger goal of providing worldwide access to biodiversity data. GBIF's goal is to make "the world's scientific biodiversity data freely available to all [**openness**]."[19] The fundamental motivation for GBIF is to enable access to a vast amount of biodiversity data housed in databases distributed in

numerous countries and institutions. By bringing all these data into one interoperable network, and producing a registry of biodiversity information resources, GBIF will produce systematic visibility and traceability of data resources [**transparency**].

Formal responsibilities of different participants involved in the task of building GBIF's organisation and legal relationships have been put forth in GBIF's Memorandum of Understanding. GBIF's Secretariat is responsible for carrying out work programmes that are approved by the Governing Board, which consists of representatives of GBIF's Participants. This structure enables GBIF to have a legal identity as an international body, manage financial contributions and work programmes, while drawing upon efforts and resources from Participants. In his reflection on the establishment of GBIF submitted to the OECD, Eric James attests: "The way in which these legal requirements are met may be the most important factor determining the structure of the organisation that is created."[20] The establishment of GBIF's activities occurred in and through contact with existing scientific and political bodies to maintain and establish professional codes, gain consensus about scientific outcomes, and negotiate with government representatives about GBIF's larger social and economic roles [**professionalism**]. The review will evaluate GBIF's progress toward data availability and interoperability, its responsiveness to user needs, and the professionalism of the Secretariat.

Participants will provide stable gateways, or "nodes," to databases that contain primary or meta-level biodiversity data. These nodes must provide documentation and metadata about the data in the databases, vouch for data **quality**, ensure data authenticity and security. GBIF will help develop standards for database **interoperability** through one of its 4 work programmes, Data Access and Database Interoperability (DADI). GBIF aims to develop an interoperable network of distributed databases by coordinating and leveraging existing national and international programs and projects, which allows for **operational efficiency** and more cost-effective basis for making biodiversity data freely and easily available to a heterogeneous user community.

The databases and the data accessed through GBIF are in most cases owned and developed by other organisations and thus will not entail any assertion of IPRs by GBIF itself [**property**]. GBIF aims to provide best practices on how to deal with IPRs, particularly since it will be drawing from databases hosted by different institutions and countries with different legal frameworks, with a view to promoting open access and sharing to the maximum extent possible.[21] GBIF also asserts in its MOU that biodiversity data will be properly used and acknowledged by its participants [**legality**]. Further, its efforts do not conflict with the Clearing House Mechanism, and they abide by the Global Taxonomic Initiative of the Convention on Biological Diversity concerning the proper and equitable use of biodiversity data and the resources to which they refer.

During the establishment of GBIF, the OECD provided the forum to assess the level of support for this new scientific collaboration, to bring together related proposals and to develop detailed plans that could then be taken up by interested countries. GBIF will have a third-year review of the effectiveness of its MOU, its scientific efforts and the "transparency of its dealings with politically sensitive issues"[22] [**accountability**].

5. Data Access Management: Five Domains

Efficient data access can only take place with the proper administration and organization of different management domains within data access regimes. These domains include technological, institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural considerations (see Figure 1). These domains provide a framework for locating and analyzing where improvements to data access and sharing can be made.

The five domains differ in character across the traditions and practices of specific scientific disciplines, e.g., astrophysics, biology. Thus, data access regimes may vary in significant ways. There is no single model for how data access should take place. The implementation of the core principle of open availability, however, requires a systematic approach that recognizes the necessity of implementing improvements across the interdependent management domains. This approach also requires the involvement of actors from various levels: governments, funding agencies, and research institutions and professional and scholarly societies, as well as individual scientists themselves.

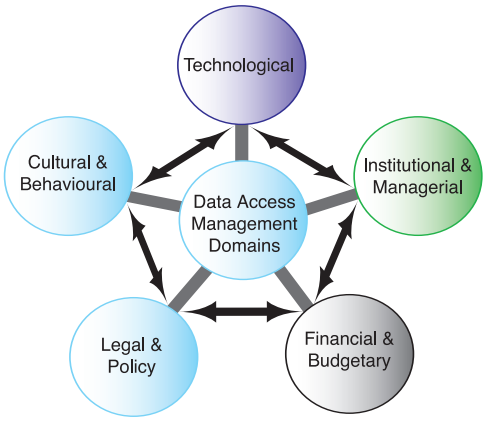

Figure 1. Components of a Data Access Regime

5.1 Technological domain: *Broad access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls.*

A technical infrastructure that supports user needs is necessary to derive maximum benefits from data access and sharing. This infrastructure must be robust enough for long term use and, when appropriate, for diverse uses. It also must be flexible enough to respond to the continuous and rapid changes in scientific research and technology. While there are many technical issues to be resolved to take full advantage of past, current and future investments in ICT infrastructure, the main barriers to effective data access and sharing are no longer technical, but are institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural.

Data Preparation and Metadata: ICPSR

In 1995, the Inter-University Consortium for Political and Social Research (ICPSR) initiated the development of the Data Documentation Initiative (DDI), an international criterion and methodology for the content, presentation, transport, and preservation of metadata about datasets in the social and behavioural sciences. DDI, which is in XML format, helps enhance users' ability to acquire and use data while it assists producers' in packaging and disseminating them. After a period of beta-testing with participating international organisations, DDI is now in use by a number of organisations, including Networked Social Science Tools and Resources (NESSTAR), Health Canada, and ICPSR. ICPSR continues to assist data producers in preparing their data through its "Guide to Social Science Data Preparation and Archiving," a guide with broad appeal for individuals and organisations searching for easy and effective ways to technically manage and prepare data so that they can be easily and effectively placed into network environments.[23]

Technical operating principles for data access regimes include **interoperability** (of protocol and software to ensure the access and usability and multiple use of the data); and **quality** (including technical components of **authenticity**, **integrity**, and **security**) of data.

5.2 Institutional and managerial domain: *While the core open access principle applies to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers.*

Because scientific data have many different characteristics and uses, there is no monolithic institutional and management approach that can be applied universally.[24] Key characteristics of data production and use include whether the data are (1) government-generated or generated at a research institution using public funds; (2) useful only within the discipline or across many disciplines; (3) useful over the very long term or only within short-term horizons; (4) have public-policy implications; or, (5) have significant broader economic and social value, among other factors.

Institutional and managerial operating principles for data access regimes include **transparency** (systematic visibility of the data source); **responsibility** (explicit formal institutional rules on data management); and **accountability** (rendering public account for the performance of data access regimes).

Negotiated collaborations: CERN

The European Organisation for Nuclear Research, CERN, is one of the world's largest scientific laboratories, presently financed by twenty European countries. CERN overtly subscribes to the core principle and premises outlined in this report, but leaves it to individual 'collaborations' of scientists to devise experiment-specific regulations to ensure compliance. Negotiations between different collaborations are necessary to enable data sharing, including agreement on definitions and standards. The type of data produced and the method of processing used will play a large part in deciding upon the most effective management model to adopt. This flexibility of management approach is a key factor in the data sharing environment at CERN.

5.3 Financial and budgetary domain: *Scientific data infrastructure requires continued, and dedicated, budgetary planning and appropriate financial support. The use of research data cannot be maximized if access, management and preservation costs are an add-on or after-thought in research projects.*

In many areas of public research, there are indications of discrepancies between the funding of the specific research itself and the related data-management requirements (which do not necessarily benefit the individual scientist, but which are necessary for data reuse). Generally, research organizations fund the former well, but pay scant attention to the latter. In the digital environment, scientific data sets must be viewed as a key element of the broader research infrastructure and as an investment in the future capacity to innovate and solve pressing problems. Adequate support is essential for data-management functions, such as the development of sufficient explanatory documentation for each data set (i.e., metadata), conversion of old formats onto new media, adaptation to new standards, and long-term preservation, archiving, and maintenance.

Budgetary operating principles for data access regimes include **operational efficiency** (maximizing the return on investment by promoting re-use of data, and providing proper documentation, specialists, and effective data management facilities).

Funding schemes "on a rolling basis:" the European Bioinformatics Institute (EBI)

The official mission of the EBI is to ensure that the growing body of information from molecular biology and genomic research is placed in the public domain and is freely accessible to the scientific community in ways that promote scientific progress. Like other scientific bodies, the EBI has a major problem in the funding for its building, maintaining and making available databases and information services even though they represent only a small fraction of the total research costs. The key issue is that funding for data sharing infrastructures needs to be constructed "on a rolling" or on-going basis to maintain effective data management. These funding requirements are very different from the funding schedules of research, which are usually project oriented. These differences in budgeting constitute the main threat to the EBI's commitment to maintaining the public availability of its data.

5.4 Legal and policy domain: *National laws and international agreements directly affect data access and sharing practices, despite the fact that they are often adopted without due consideration of the impact on the sharing of publicly funded research data.*

Intellectual property laws, information policies, institutional guidelines, and contracts at the national and international levels often impose terms and conditions on data access and sharing practices. Laws and policies governing data access and sharing practices may vary among different countries, resulting in barriers to scientific cooperation and progress. Based on a recent Web survey, most of the national research organization managers who responded expected that data sharing will become a major policy issue in the next five years. This situation requires greater attention by the science policy community at all levels. In particular, restrictions on re-use of public data by the research community must be eliminated or minimised as much as possible. Research grant provisions and licensing templates for promoting open access and unrestricted re-use of public research data already exist, but have not yet been broadly adopted.

Policy interconnections: functional MRI and the Institutional Review Boards

The functional Magnetic Resonance Imaging Data Center's (fMRIDC) principal endeavour is to promote data sharing in brain mapping. The Western tradition of informed consent in bio-medicine operates according to the principle that the 'most specific consent is the best consent.' When data are to be gathered for submission to databases, the specificity of consent may run counter to the goals of meta-analysis or re-analysis by third parties, to investigate issues different from those for which the data was originally gathered. The creation of infrastructures for data sharing, therefore, has to conform to the rules of regulatory bodies, such as institutional review boards (IRBs), whose approval must be obtained to share data. As such, these bodies function as gatekeepers to the circulation of data. International coordination may also be necessary. Researchers submitting or requesting data across national boundaries may find it especially difficult to act in accordance with the various ethical guidelines that exist in different countries. The fMRIDC has been hesitant to accept data from non-US settings because of concerns regarding IRB compliance.

Legal and policy operating principles for data access regimes include **property** (balance intellectual property rights of investigator and institution versus public good); and **legality** (lawful data management, respecting national security, privacy and trade secrets).

5.5 Cultural and behavioural domain: *Appropriate reward structures are a necessary component for promoting data access and sharing practices. These apply to both those who produce and those who manage research data.*

Although formal policy frameworks and regulations are necessary to make research data publicly available, they need to be supplemented by appropriate community-based norms and incentives for researchers to share and provide access to their data and for appropriate recognition of their data-related work. In many cases, there is a general lack of reward structures and mechanisms to promote open access to, and sharing of, data from public research.

Cultural and behavioural operating principles for data access regimes include quality (trust that data are what they purport to be); professionalism (build on codes of conduct and ethics of the scientific community); flexibility (there is no single model on how data access must be provided.)

Incentives: the Protein Data Bank

To publish in scientific journals, U.S. scientists involved in the field of crystallography must deposit their data in the Protein Data Bank (PDB) and acquire an accession number. "By requiring everyone to submit data, the community is assured of having the most up to date information possible. Now, increasingly, under our regime, a lot of [data] depositors have come to realize that the practice that we use has some advantages for them in that we check things and we find errors and inconsistencies. That actually improves the quality of the product they produce."[25]

6. Possible Action Areas

Our findings from the case studies and from other research indicate a number of action areas by the different actors involved in making possible open access to, and sharing of, publicly funded research data. In this section we recommend possible action areas for the OECD and national governments.

OECD

As an international organization with credibility and stature in the science policy arena, the OECD, through the CSTP, can play a crucial role in promoting access to, and sharing of, data from publicly funded research. Central to this role is the gathering and sharing of information on data related activities and policies. At the international level, only a small handful of organizations have undertaken to do this, usually in the context of a specific discipline or research program. The recent, and vast, expansion of research data assets and the trend towards issue-based, interdisciplinary research, however, suggests that all countries and all fields of science stand to benefit from greater attention and an organized and coordinated approach to effective policy actions

1. **The OECD should put the issues of data access and sharing on the agenda of the next Ministerial meeting.** ICT advances have created the ability to transform science. New tools allow researchers to find data in seconds that would have taken months just a few years ago. Effective data access and sharing requires a comprehensive policy approach for implementation by public research institutions. Monitoring progress and devoting attention to the public research data issues and activities would assist decision-makers and research support agencies in developing appropriate policies and allocating resources

Areas in Conjunction with Relevant Member Country Research Organizations

2. **The OECD should consider conducting or coordinating a study to survey national laws and policies that affect data access and sharing practices.** This relatively simple undertaking could determine what policies exist, how accessible they are, and result in listing of the web sites where these policies are posted. This study would be of considerable benefit to science policy-makers, research administrators, and information resource managers in all countries, both within OECD and beyond. The study could look at the feasibility of developing a central and easily accessible repository of national laws and policies that affect data access and sharing practices. Such a compilation does not currently exist, and could be useful to facilitate international research collaborations.
3. T**he OECD should consider conducting or coordinating a study to compile model licensing agreements and templates for access to and sharing of publicly funded data.** Depending on the context, numerous factors need to be considered in data access and sharing arrangements. Nevertheless, many contractual models already exist that have been developed by research funding organisations, research program managers, university administrators, librarians, and others. The OECD, as a global organization, is ideally suited to span national domains where examples do exist, and thereby bring an international perspective. The study could compile and review existing agreements and models to find exemplary approaches. Having readily available models on hand would be of considerable benefit to researchers, universities, and research institutions, as well as data centers and archives, and could facilitate international research collaboration

Areas for Further Examination

4. **Governments within the OECD should expand their policy framework of research data access and sharing to include data produced from a mixture of public and private funds.** Collaborative public/private research projects, and the resulting data, have their own unique set of characteristics and issues. As more national governments promote public-private partnerships in research, these issues will be of increasing importance to both public researchers and the companies that are involved. A further examination of the state of data sharing and access in these types of research arrangements needs to be made to develop sound science policy guidance.
5. **The OECD should consider examinations of research data access and sharing to include issues of interacting with developing countries.** The increase of participation in the research enterprise benefits the global science system and innovation. Providing developing countries with access to data from publicly funded research increases their participation in science. Further, as United Nations Education, Scientific and Cultural Organization (UNESCO), the International Council of Scientific Unions (ICSU), private foundations, and other organizations have emphasized, access to scientific knowledge by developing countries is vital to the progress of the entire world. This access is particularly important in the context of global issues such as population health, environmental change, and food production. Of course, open access to data from publicly funded research in developed countries can provide a valuable resource for economic development, education, and scientific capacity building. Many efforts are already underway to improve access for researchers in developing countries (e.g. providing free or below costs access to data and scientific information) as well as establishing optimal data regimes for developing countries to share their data (e.g. addressing issues of data repatriation). A systematic examination of barriers and best practices would provide both a picture of the current situation and a set of guidelines for further action.
6. **The OECD should promote further research, including a comprehensive economic analysis of existing data access regimes, at both the national and research projectd or program levels**. To date, no one has yet undertaken a comprehensive, economic analysis of different data access regimes. Several key issues have not been closely examined, including the relative costs of providing data openly, the impact of cost recovery on the use of those data, and the positive externalities and network effects from providing open access to publicly funded research data. The OECD should consider conducting this type of analysis or encouraging member country research organizations to fund such studies.

National Governments

Although the OECD, UNESCO, ICSU, and other international bodies can play a role in improving the current situation regarding research data access and sharing, it is at the national level that many important decisions and actions must be taken. National governments provide the resources for making data accessible, establish the overall policies for data management, regulate matters such as confidentiality and privacy, and determine restrictions based on national security. Most importantly, it is national governments that are responsible for the major research support and funding organizations, and it is here that many of the managerial aspects of data sharing need to be addressed.

The national governments of OECD countries should consider:

1. **Adopting, and effectively implementing, the principle that data produced from publicly funded research should be openly available to the maximum extent possible.** The public investments made in research data collection can only be maximized if the data are preserved, managed, and made accessible. This requires coordinated attention by governments at all levels, and adequate policy and financial support. The starting point for these actions, however, is the affirmation that data collected using public funds should be openly accessible to all.
2. **Encouraging their research funding agencies and major data producing departments to work together to find ways to enhance access to statistical data, such as census materials and surveys.** Many countries have taken steps to facilitate access to census and survey materials by developing catalogues, user-friendly repositories, off-site research facilities, training programs, and regulatory frameworks for providing appropriately guarded access to confidential information. Such steps have proven enormously effective in maximizing the use of national surveys and producing insights into the functions of economies and societies.
3. **Adopting free access, or marginal cost pricing, policies for the dissemination of research-useful data produced by government departments and agencies.** The use of information collected through public funding should be freely accessible for research purposes. This maximizes the use of such information for public policy and public knowledge development.
4. **Analyzing, assessing and monitoring policies, programs, and management practices related to data access and sharing policies within their national research and research funding organizations.** This information would be useful to national governments so that they may assess the implementation of the previous three considerations. The resources, support programs, policies, and regulations related to research data sharing are, in large part, developed and implemented by research funding organizations. The operations of these organizations play a crucial role in determining the degree to which data are made accessible and shared between researchers. Many organizations, such as NSF and NIH in the United States, Social Sciences and Humanities Research Council in Canada, and the European Science Foundation are now developing, or have developed, policies, regulations and support programs that promote data sharing. Issues such as establishing protocols for the collection and release of confidential information, developing technical infrastructure, agreeing on metadata standards, requiring data preservation strategies within individual research projects, and including data management costs as eligible expenditures in grant applications have been dealt with by one or more of these agencies. It would benefit the global scientific community if decision-makers within national governments had a clear understanding of where their respective agencies stood in relation to those in other countries.

7. Conclusion

Improving access to and sharing of publicly funded research data is an issue that touches on all aspects of the research enterprise and the development of knowledge, and involves all participants in the conduct of research. For the individual researcher, the sharing of data, particularly prior to publication[26], can be burdensome, time consuming, and unrewarding if the necessary measures are not taken to provide funding, facilities, and a social context that emphasises its value to the research community and to society.

Advances in ICTs, the internationalisation of science, and the trend toward issue-based research hold great potential for the advancement of knowledge and for the benefit of all people. This potential will not be fully realized unless all of the major elements of data access regimes identified in this report are properly developed. To do so will take considerable discussion, understanding, and commitment on the part of all those involved in research, particularly at the policy level.

Agreement among OECD governments on a set of general principles to shape specific data access regimes, as well as adoption of the recommendations set forth above, would be enabling for scientists, empowering for citizens, and provide an important contribution to fulfill the promises of e-science.

---

[1] In this report, we define "access to data" as the act of making the data available for use by others; by "sharing" we mean a researcher allowing one or more other individuals to use data, typically with the implicit, if not explicit assumption that it is on a reciprocal basis. The sharing of data involves providing specific access, whereas the act of providing access by itself does not necessarily involve any sharing arrangement. Data sharing focuses on data exchanges between individual researchers rather than institutions, while access may be provided at any level. Sharing also reflects the cooperative norms of public science as practiced within many disciplines by many researchers in OECD countries. We define data as in the U.S. National Institutes of Health definition of final research data: “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings”.

[2] See http://www.mrc.ac.uk/index/strategy/strategy-science_strategy/strategy-strategic_implementation/strategy-data_sharing/strategy-data_sharing_policy-link

[3] CODATA, the interdisciplinary Committee on Data for Science and Technology of ICSU, is currently examining barriers to data access and sharing that are particular to developing countries. CODATA, however, does not normally examine issues related to social science and humanities research. Related to issues of national security, see “NAS Censors Report on Agricultural Threats,” *Science* 20, p. 1973-1975, on the several scenarios that were left out of a public report of the U.S. National Academy of Sciences.

[4] See NSF 2002 Science and Engineering Indicators, http://www.nsf.gov/sbe/srs/seind02/start.htm

[5] Gary Stix (2001), “Triumph of Light,” *Scientific American*, January, available at http://www.sciam.com/2001/0101issue/0101stix.html

[6] Examples range from genetic sequence and protein structure data in bioinformatics, to various types of brain imagery in neuroscience, to sky surveys and virtual observatories in astronomy, and geospatial data such as Global Spatial Data Infrastructure.

[7] Examples include combining data from multiple data sources to gain a greater statistical power to resolve hypotheses (see the Biomedical Informatics Research Network, http://www.nbirn.net); and obtaining real-time global measurement on environmental observations.

[8] John Taylor, Director General of (UK) Research Councils (UK), www.research-councils.ac.uk/escience/. “E-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientist. . . . Besides information stored in Web pages,

scientists will need easy access to remote facilities, to computer – either as dedicated Teraflop computers or cheap collections of PCs – and to information stored in dedicated databases. The Grid is architecture to bring all these issues together." See also Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Report on Cyberinfrastructure, http://www.cise.nsf.gov/evnt/reports/atkins_annc_020303.htm .

[9] Rita Colwell (2002), "A Global Thirst for Safe Water: The Case of Cholera," Abel Wolman Lecture at the National Academy of Sciences, January 25, 2002, available at http://www7.nationalacademies.org/wstb/2002_Wolman_Lecture.pdf. Other examples of the impact of access to and sharing of international data in the control and elimination of worldwide diseases include the World Health Organisation's network of collaboration centres. In the worldwide programme of epidemiological surveillance of influenza, these receive epidemiological information on outbreaks of influenza from national institutions throughout the world. They also receive new strains of the virus for characterization and give advice as to their possible use in vaccine preparation. The centres then distribute the necessary reagents, antigens and anti-sera to national laboratories, and high-yielding recombinant viruses for to vaccine producers. See http://whqlily.who.int/general_infos.asp.

[10] For more benefits of data sharing, see National Academy Press (1985), *Sharing Research Data*, available at http://books.nap.edu/catalog/2033.html

[11] Peter Weiss (forthcoming 2003) presentation in Proceedings of the Symposium on the Role of Scientific and Technical Data in the Public Domain, National Academies Press. See also, European Union Green Paper (1998), "Public Sector Information: A Key Resource for Europe," COM 585, and PIRA International, "Commercial Exploitation of Europe's Public Sector Information, Final Report for the European Commission (2000),"Directorate General for the Information Society," which provide similar comparisons of such policies in other information sectors.

[12] Ronald C. McMahon (1996), "Cost Recovery and Statistics Canada," in Government Information in Canada, Volume 2, number 4 (spring 1996), retrieved from http://www.usask.ca/library/gic/v2n4/mcmahon/mcmahon.html, February 2003

[13] Studies include: National Research Council (1997), Bits of Power: *Issues in Global Access to Scientific Data*, National Academy Press, Washington, D.C.; National Research Council (1999); and *A Question of Balance: Private Rights and The Public Interest in Scientific and Technical Databases*, National Academy Press, Washington, D.C.; Stephen Hilgartner (1996), "Access to Data and Intellectual Property: Scientific Exchange in Genome Research" in *Intellectual Property Rights and the Dissemination of Research Tools in Molecular Biology: Summary of a Workshop held at the National Academy of Science, February 15-16, 1996*; and National Research Council (1995),*On the Full and Open Exchange of Scientific Data*, National Academy Press, Washington, D.C.; and National Research Council (2002), C*ommunity Standards for Sharing Publication-Related Data and Materials*, National Academy Press, Washington, D.C. The European Bioinformatics Institute, the Global Change Program, the Global Biodiversity Information Facility, the European Social Survey, the International Union of Crystallography, the international Ocean Drilling Program; The European Organization for Nuclear Research, otherwise known as CERN, provide good examples of research programmes with effective data policies. Funding agency statements include: NSF at http://www.nsf.gov/sbe/ses/common/archive.htm) and "NIH Draft Statement on Sharing Research Data" at http://grants2.nih.gov/grants/policy/data_sharing/. Sites that discuss how to develop a data policy include Smithsonian Environmental Research Center at http://www.serc.si.edu/datamgmnt/policy1.htm and the Ecological Sciences Network at www.esnet.edu. The policy of the Long Term Ecological Research (LTER) network is at http://www.lternet.edu/data/netpolicy.html.

[14] For more on the Role of Governments in the Digital Age, see Stiglitz, Orzag and Orzag at http://www.ccianet.org/govt_comp.php3. In particular note the following three:

Principle 1: Providing public data and information is a proper governmental role

Principle 2: Improving the efficiency with which governmental services are provided is a proper governmental role

Principle 3: The support of basic research is a proper governmental role

[15] As one researcher put it, "Incentives for data sharing need to be offered that offset the investigators' loss of control over their databases. Usually, this is some form of added scientific value. By sharing data, an

investigator may gain access to more data or other tools. Ultimately, there has to be a procedural framework that makes sharing sensible, efficient, and value-added. If all those pieces are in place, fewer external or coercive forces are needed to convince researchers to share." From minutes from an NIMH meeting, see Paul Wouters, Data Sharing Policies, 10 June 2002. Networked Research and Digital Information, NIWI-KNAW on http://dataaccess.ucsd.edu

[16] In economics, a good is considered a "public good" if it is "non-rivalrous" and "non-excludable." The former means that the marginal costs of providing the good to an additional person are zero. The latter means that once the good is produced, the producer cannot exclude others from benefiting from it. See, Inge Kaul, Isabelle Grunberg, and Marc Stern (1999), "Defining Global Public Goods," in *Global Public Goods: International Cooperation in the 21st Century*, eds. Both publicly funded basic research and the data produced from it and disseminated on digital networks are non-rivalrous. They are not purely excludable, however, although their excludability, especially for other researchers, is neither economically efficient nor desirable as a matter of public policy, absent countervailing and superseding reasons to the contrary.

[17] These operating principles evolved from the document produced by Hans Franken, Access to Publicly Financed Research, Conference Conclusion. Global Research Village III Amsterdam 2000. For other principles on data access and sharing see http://www.codata.org/data_access/principles.html . Examples of successful guidelines based on a systematic set of principles are the OECD Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data (1980) and the Principles and Guidelines for the Sharing of Biomedical Research Resources (1999) from the US National Institutes of Health (NIH) and the OECD Guidelines for Security of Information Systems and Networks (2002).

[18] Examples include *National security*: Data sets from some oceanographic or geological surveys may be (partly) classified and not accessible; Privacy: Data from human subjects are vulnerable to breaches of confidentiality and privacy and therefore should only be obtained by fair and lawful means, with knowledge or consent of the data subjects; and *Trade secrets*: Data potentially relevant to prospective patenting or commercial opportunities may contain (partly) confidential information.

[19] See www.gbif.net

[20] Eric James, "Establishing International Scientific Collaborations: Lessons Learned from the Global Biodiversity Information Facility," submitted to Sixth Meeting of the OECD Global Science Forum, available at http://www.oecd.org/pdf/M00027000/M00027203.pdf

[21] Research on issues of IPR, particularly for natural history museums, is being conducted by European Natural History Specimen Information Network, see. http://www.nhm.ac.uk/science/rco/enhsin/details.html and "Beset with pitfalls-specimens and databases, intellectual property and copyright," Simon J. Owens and Alyson Prior, from the 2000 meeting of the Taxonomic Databases Working Group, November, 2000; Senckenberg Museum, Frankfurt, available at http://www.tdwg.org/tdwg2000/ipr.htm.

[22] Eric James, "Establishing International Scientific Collaborations: Lessons Learned from the Global Biodiversity Information Facility," submitted to Sixth Meeting of the OECD Global Science Forum, Section 10.

[23] For more information on ICPSR and DDI, http://www.icpsr.umich.edu/DDI/ORG/index.html. For ICPSR's Guide, see www.icpsr.umich.edu/ACCESS/dpm.html. For information on the importance and development of DDI, see "Providing Global Access to Distributed Data through Metadata Standardisation -- The Parallel Stories of NESSATAR and DDI", submitted by the Norwegian Social Science Data Services to the Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata, Geneva, Switzerland, 22-24 September 1999, at http://www.nesstar.org/papers/GlobalAccess.html.

[24] For example, mathematics presents a special feature in that published material never becomes obsolete, so that the data needed by the working mathematician is ideally the full collection of published papers, past and present. With the development of internet access, this is not an impossible objective. New papers are almost always produced in electronic form, and therefore could be stored and accessed. The amount of past literature to be scanned and digitized is estimated to be around 50 million pages. Under the umbrella of the International Mathematical Union, an attempt is made to coordinate national efforts to insure permanent accessibility at a reasonable cost for the users to both new and digitized papers. Without this, research will be limited to rich parts of the world, where libraries can be heavily funded. See http://www.mathematik.uni-bielefeld.de/~rehmann/DML/ and http://www.library.cornell.edu/dmlib/.

[25] Berman, Helen. Director, Protein Data Bank. Personal communication.
[26] Rapid Data Release Policy: "Ever since the 1996 Bermuda Principles provided guidelines on the rapid release of data from large-scale sequencing projects, access to the pre-publication sequence data that has been made freely available in public nucleotide sequence databases has accelerated biomedical research. However, in 2002, it became clear that new strategies and other advances in large-scale DNA sequencing necessitated a re-examination
and updating of the data release policies originally developed to implement the Bermuda Principles for pre-publication sequence data. At its February 10-11, 2003 meeting, the National Advisory Council for Human Genome Research (NACHGR), the main advisory group to the National Human Genome Research Institute (NHGRI) on genetics and genomic research, discussed the subject of pre-publication release of large-scale sequencing data. NACHGR approved a draft policy that would reaffirm and extend the rapid data release policies developed to implement the 1996 Bermuda Principles, and recommended that NHGRI publicize the draft policy statement for the purpose of obtaining comment from the scientific community." (http://www.genome.gov/page.cfm?pageID=10506376). For a reaffirmation and extensions of the NHGRI rapid data release policy, see http://www.genome.gov/page.cfm?pageID=10506537. For community discussion see Sacrifice for the greater good? Nature 421, 875 (2003), and Draft guidelines ease restrictions on use of genome sequence data, Nature 421, 877-878 (2003).

10.3 Rapport final de la Consultation sur les archives nationales de données - Juin 2002 (sans les annexes)

Consultation sur les archives nationales de données

Rapport final

Infrastructure pour l'archivage et l'accessibilité des données de recherche du Canada

Présenté par le groupe de travail – Consultation sur les archives nationales de données (CAND)
au Conseil de recherches en sciences humaines du Canada et
à l'archiviste national du Canada

Résumé

En octobre 2000, le Conseil de recherches en sciences humaines et l'archiviste national du Canada ont constitué un groupe de travail comprenant des chercheurs et des archivistes et leur ont demandé d'évaluer les besoins en ce qui a trait à un système national d'accessibilité, de conservation et de gestion de données de recherche. Ayant compilé de nombreuses preuves à l'appui d'un tel service de soutien à la création de connaissances par la communauté des chercheurs œuvrant dans le domaine des sciences humaines, le groupe de travail présente maintenant ses recommandations concernant la création d'un nouveau service national d'archivage de données de recherche. Ce service aura trois fonctions de base :

- Conserver les données de recherche compilées par les chercheurs et les organismes gouvernementaux, les entreprises de sondage et autres organismes, qui peuvent être utilisées par les chercheurs pour créer de nouvelles connaissances;
- Gérer les données déposées et en assurer la qualité; sélectionner les données qui seront conservées, élaborer et appliquer des normes relatives aux métadonnées, à l'authentification et à la sécurité et diffuser les données dans l'ensemble des technologies;
- Fournir l'accès aux données de recherche, y compris aux systèmes de sortie Web, aux services de catalogage, aux ententes avec les utilisateurs et les dépositaires afin de protéger la confidentialité et les droits de propriété intellectuelle et aux connexions aux autres dépositaires de données dans le monde entier.

De plus, le groupe de travail recommande qu'un nouveau réseau national d'archives de données de recherche intègre certaines autres fonctions, y compris organiser des cours de formation spécialisée dans les techniques de maniement des données, représenter les intérêts canadiens dans le développement de normes de données internationales, promouvoir l'échange des données comme pratique exemplaire en recherche, effectuer des recherches dans le domaine des sciences de l'information et de l'archivistique, et agir en tant qu'organisme central de coordination pour un réseau de services de données, dans les instituts de recherche canadiens.

L'information numérique compilée aux fins de la recherche joue un rôle de plus en plus important dans l'économie actuelle fondée sur le savoir. À bien des égards, les données sont le moteur de l'innovation et de notre capacité à résoudre des problèmes sociaux et économiques complexes. Bien que des milliards de dollars soient dépensés chaque année pour la collecte de données, le Canada ne possède pas l'infrastructure requise pour en assurer la conservation et l'accès par le public, ce qui limite le rendement du capital investi dans la recherche et mine la bonne gestion publique.

De nombreuses composantes de base requises pour la création des archives nationales de données de recherche existent déjà, notamment, les services de traitement des données des universités, les réseaux de transmission à haute vitesse, les directives juridiques et éthiques et les cadres de travail, les instituts partenaires potentiels, divers dépositaires de données et des initiatives relatives à un portail d'accès ainsi qu'une communauté de chercheurs qui produisent effectivement des données. L'élément manquant est un service de conservation, de coordination et de gestion des données.

Presque tous les pays développés ont reconnu le besoin d'un service national d'archives de données de recherche et certains pays ont plus d'une génération d'expérience dans l'exploitation de tels services. Le Canada peut bénéficier de cette expérience tout en développant un service d'archives de données de recherche compatible avec son contexte institutionnel et culturel unique.

Nous possédons maintenant la capacité technologique et le savoir-faire requis pour créer un « système de confiance » en vue de fournir aux Canadiens et aux Canadiennes un service accessible et étendu qui permettrait aux chercheurs de localiser, de demander, d'extraire et d'utiliser les ressources en données de façon simple, transparente et économique tout en protégeant la vie privée, la confidentialité et les droits de propriété intellectuelle des intéressés. La Fondation canadienne pour l'innovation pourrait financer les coûts liés à l'élaboration de l'infrastructure de ce service. On estime que les coûts de fonctionnement annuels d'une installation et d'un réseau étendu sont d'environ 3 millions de dollars.

Le groupe de travail présente trois options pour ce qui est de la création d'un réseau national d'archives de données de recherche :

1. Créer, par le truchement de la législation fédérale, un réseau national d'archives de données de recherche en tant que version modifiée d'un organisme distinct constitué en vertu d'une loi. Cette option serait idéale pour la mise en place d'un service complet, d'un organisme fiable comprenant une installation centrale de conservation et de gestion des données et un ensemble de nœuds de services d'accès localisés dans des instituts de recherche. Le service profite pleinement de l'infrastructure de recherche existante, est stable à long terme et constitue une connexion directe pour les utilisateurs et les producteurs de données de recherche, et est apte à représenter les intérêts du Canada dans le développement des normes internationales de données.
2. Créer un réseau national d'archives de données de recherche sous la gouverne du Conseil de recherches en sciences humaines. Cette stratégie englobe les caractéristiques du premier modèle, mais ne requiert pas de dispositions législatives. Elle prend avantage de la connexion directe et immédiate avec les chercheurs, de l'obligation reconnue de rendre compte et des structures de financement.
3. Créer un organisme de service spécial au sein des Archives nationales du Canada. Par la création d'une division autonome au sein des Archives nationales, cette option prend avantage de l'infrastructure et de l'expertise existantes en matière d'archivage. Cette stratégie n'a pas été retenue par les autres pays parce que la mission fondamentale d'un service d'archivage national des données de recherche est essentiellement différente. Néanmoins, à titre d'organisme de service spécial, le service pourrait possiblement être stable et en mesure de développer un système fiable d'accès aux données, de gestion et de conservation de celles-ci.

Comme prochaine étape, le groupe de travail recommande que le CRSH et l'archiviste national constituent un comité directeur qui sera chargé de sélectionner la stratégie appropriée pour la mise en place d'un réseau national d'archives de données de recherche ou d'un service d'archivage des données de recherche et de définir plus précisément les caractéristiques et les exigences en matière de financement d'un tel service et de promouvoir sa mise en place.

Table des matières

Sections

1. Introduction

2. En quoi consiste l'archivage des données de recherche?

3. Besoins en matière d'archives nationales de données de recherche

4. Composantes de base des archives nationales de données de recherche

5. Vers un modèle institutionnel : leçons tirées de la scène internationale

6. Principes fondamentaux et hypothèses

7. Options pour le Canada

8. Recommandations concernant la mise en place d'un réseau national d'archives de données de recherche

9. Coûts d'un réseau national d'archives de données de recherche

10. Prochaines étapes

Infrastructure pour l'archivage et l'accessibilité des données de recherche du Canada

1. Introduction

Au Canada, le réseau d'instituts et d'organismes qui constitue l'infrastructure de soutien dont dépend l'économie fondée sur le savoir souffre actuellement de lacunes importantes. Le Canada ne possède aucun organisme national pour conserver et cataloguer les données et fournir un accès systématique, efficace et opportun aux données de recherche. Ces renseignements numériques permettent aux chercheurs de corroborer les connaissances existantes, de reproduire les essais, de vérifier les résultats de la recherche, d'examiner les données et de créer de nouvelles connaissances. L'accès efficace aux données de recherche et leur utilisation peuvent jouer un rôle central pour ce qui est de la capacité d'innovation du Canada. Toutefois, l'infrastructure requise doit être mise en place.

En octobre 2000, le Conseil de recherches en sciences humaines et l'Archiviste national du Canada ont donné à un groupe de travail composé de chercheurs et d'archivistes le mandat de consulter la communauté des chercheurs et archivistes et d'évaluer le besoin d'un service national ou d'une fonction d'archivage des données. Au terme de l'évaluation et de la compilation d'éléments probants confirmant le besoin d'un tel service, le groupe de travail a étudié les archives de données de recherche d'autres pays et examiné les voies possibles en vue de créer une installation de recherche de base au Canada.

Le groupe de travail recommande maintenant la création d'un organisme canadien en vue de combler cette lacune dans l'infrastructure de l'économie canadienne basée sur le savoir – la création et le soutien à long terme d'un système d'archives nationales de données de recherche.

De nos jours, presque toute la recherche est effectuée dans un environnement numérique. Des bases de données statistiques complexes à multicouches, des cartes et images numériques et des textes codés constituent maintenant les outils usuels des chercheurs. Bien que ces ressources aient donné une envergure spectaculaire à la recherche et accru son efficacité, les structures institutionnelles exigées pour la conservation et la gestion des données et l'accès aux données numériques ne sont pas à la hauteur. Cette situation mine la capacité d'innovation des chercheurs canadiens et met en péril des données de recherche précieuses valant des dizaines de millions de dollars.

L'édification d'une société axée sur le savoir, la stimulation de l'innovation et l'examen des problèmes urgents et complexes sur les plans social, politique et économique reposent en grande partie sur la découverte de connaissances par le biais de la recherche. Pour être réceptifs et efficaces, tout en incorporant des perspectives multiples, les chercheurs doivent avoir accès à un large éventail de données de recherche et avoir la possibilité de les partager. Une infrastructure est nécessaire pour atteindre cet objectif. Aujourd'hui, de nombreux éléments sont en place, tels les bibliothèques de recherche et les services informatiques des universités, les conseils de soutien à la recherche, les réseaux de transmission de données à haute vitesse – mais il manque un élément essentiel, soit une installation de stockage, de distribution et de conservation des données de recherche.

La saine gestion publique exige que l'investissement public dans les données de recherche génère un rendement maximal du capital investi. Afin de maximiser le rendement, les données de recherche devraient être utilisées autant de fois que possible dans autant de contextes différents et cela ne peut se produire que si nous mettons en place une infrastructure efficace de données de recherche. Le coût de l'inaction met nos investissements en sciences en péril et mine également une des responsabilités essentielles du gouvernement.

2. En quoi consiste l'archivage des données de recherche?

À la différence de nombreuses formes d'archivage traditionnelles, l'archivage des données de recherche consiste non seulement à tenir des registres à des fins juridiques, historiques ou culturelles, mais il doit satisfaire aux besoins des chercheurs qui travaillent dans l'environnement numérique actuel. La mission fondamentale des archives de données de recherche n'est pas de conserver les mémoires enregistrées d'un groupe, d'un organisme ou d'une nation, mais de fournir un service essentiel au milieu de la recherche.

Bien que les nouveaux besoins des instituts en matière de programmes numériques soient de plus en plus nombreux, le groupe de travail n'a examiné que les besoins de la communauté des chercheurs en matière d'accès aux données, de gestion et de conservation de celles-ci, en particulier les données en sciences humaines. À partir de cette perspective, le groupe de travail a défini le processus d'archivage de données comme étant la conservation et la gestion des données de recherche et l'accès du public aux renseignements numériques structurés au moyen de méthodes de recherche dans le but de créer de nouvelles connaissances. Ce processus permet d'utiliser les résultats de la recherche qui sont produits depuis l'information initiale jusqu'à la publication des résultats. Les acquisitions comprendraient les renseignements numériques fournis par les chercheurs et qui présentent de l'intérêt pour eux, sous réserve des limites des ressources financières et des protocoles de conservation élaborés par les milieux de la recherche et des archives de données de recherche eux-mêmes.

Les archives et les organismes nationaux de données de recherche d'autres pays offrent une vaste gamme de services d'accès, de conservation et de gestion à leurs milieux de recherche respectifs, y compris le stockage sur le site et hors site, l'accès à des catalogues et à des ensembles de données sur Internet, des protocoles de conservation, la création de métadonnées, la migration des données par des logiciels et du matériel informatique, la formation et l'élaboration de normes internationales. En offrant ces services, ces pays jouent un rôle actif et essentiel en tant que diffuseurs de

l'information et des connaissances.

Les Archives nationales de données m'apparaissent comme un institut canadien de confiance ayant pour mandat de conserver les données de recherche, de collaborer avec d'autres organismes gouvernementaux et non gouvernementaux pour assurer que leurs pratiques de gestion des données incorporent des normes de conservation, de travailler étroitement avec d'autres instituts canadiens chargés de préserver l'héritage du Canada et de prévenir les lacunes en matière de responsabilités, de représenter le Canada à l'échelle internationale lors des échanges de données de recherche et de l'élaboration de normes connexes, en vue de permettre l'accès à ces données, d'informer les Canadiens sur la manière d'utiliser les données de recherche, de contribuer aux nouvelles recherches en créant de nouvelles données à partir des données archivées, de participer à la protection de la vie privée dans la société canadienne compte tenu du stockage considérable d'informations numériques sur les personnes et de mener des activités de recherche et développement sur tous les aspects de la conservation des données.

—Charles Humphrey, bibliothécaire de données, University of Alberta,
Membre du Groupe de travail, Consultation sur les archives nationales de données

---

3. Besoins en matière d'archives nationales de données de recherche

Comme l'a mentionné l'un des membres du groupe de travail, une « tempête de feu » sans précédent est en train de ravager la richesse du Canada en matière de données numériques. Cet énoncé peut sembler exagéré mais, en réalité, il traduit les préoccupations profondes de nombreux archivistes, bibliothécaires et chercheurs dans le monde entier[1].

Les résultats de recherche sous forme numérique sont extrêmement fragiles, mais des quantités considérables de ces données peuvent être compilées. Aujourd'hui, nous commençons à peine à comprendre comment conserver et gérer efficacement cette information. Bien qu'il n'existe aucun raccourci pour résoudre des problèmes tels l'obsolescence des médias, la « désuétude » du système numérique, le droit d'auteur, la confidentialité, l'élaboration de normes nationales et internationales et les limites des milieux de recherche actuels, éviter ces problèmes ou les ignorer pourrait s'avérer onéreux, à long terme.

Dans la phase initiale de la Consultation sur les archives nationales de données, le groupe de travail a consulté une vaste gamme d'intervenants qui utilisent, gèrent et produisent des données de recherche relatives aux sciences humaines. L'objet du groupe de travail était d'évaluer le besoin d'un service national ou d'une fonction d'archivage de données de recherche (voir l'annexe E). Cette évaluation a permis de déceler un certain nombre de lacunes structurelles :

- Actuellement, il n'existe aucun institut national ayant le mandat de conserver, de gérer et de produire des données de recherche accessibles au public, qui ait l'envergure requise pour appuyer les efforts de la communauté canadienne des chercheurs. Les Archives nationales du Canada ne disposent pas des ressources requises à cet effet;
- Les services de recherche de données des universités n'ont ni les ressources, ni la responsabilité les habilitant à agir à titre d'archives de données de calibre national. Bien qu'ils s'efforcent de combler les lacunes qu'entraîne l'absence d'un service national d'archivage des données, les services de traitement des données des universités n'ont, en général, que le mandat de fournir aux usagers locaux l'accès aux données facilement disponibles;
- La politique du CRSH sur l'archivage des données, qui incite les chercheurs qu'il soutient à déposer leurs données dans les dépôts d'archives des services de traitement des données des universités, n'a pas atteint ses objectifs. En réalité, en 11 ans, seulement 10 ensembles de données ont été déposés dans les dépôts d'archives de données des universités figurant dans le Guide du CRSH. Bien que certains chercheurs hésitent à partager leurs données, il serait contraire à l'éthique que le CRSH applique cette politique en l'absence d'une installation qui permettrait aux chercheurs de se conformer à la réglementation;

[1] Nombreux sont les organismes qui se débattent actuellement avec les structures et politiques relatives à l'archivage des données, y compris la Library of Congress, l'Economic and Social Research Council du Royaume-Uni, les National Institutes of Health des États-Unis, les National Archives and Record Administration, et le National Research Council des États-Unis, l'International Council for Scientific and Technical Information, et l'International Council of Scientific Unions.

- Le Canada ne dispose d'aucune stratégie coordonnée pour établir des normes internationales de données de recherche dans le cas des projets de métadonnées, telle l'Initiative de documentation des données, ni d'outils pour accéder aux données, tel le projet Networked Social Science Tools and Resources (NESSTAR), et de projets collectifs d'infrastructures internationales, comme Frameworks, de l'Union européenne. De même, le Canada n'est pas représenté à l'échelle nationale à l'International Federation of Data Organizations et ne participe pas aux initiatives du Council of European Social Science Data Archives;
- L'un des problèmes primordiaux auxquels les chercheurs sont confrontés aujourd'hui est la difficulté de localiser les données relatives à la recherche qu'ils effectuent. Il n'existe aucune liste collective ou catalogue des ensembles de données que possèdent les producteurs et distributeurs de données et autres chercheurs. Il s'ensuit que les chercheurs peuvent reproduire inutilement des études coûteuses, se fonder sur de l'évidence anecdotique plutôt que sur l'évidence empirique ou utiliser des données de remplacement provenant d'autres pays. Un service national de données pourrait transférer directement aux ordinateurs personnels des chercheurs l'information concernant les sources des données ainsi que les données elles-mêmes, ce qui entraînerait une économie de temps et d'argent.

Des archives nationales de données de recherche pourraient également répondre aux besoins fondamentaux, de la façon suivante :

- Assurer l'authenticité des données de recherche, ce qui constitue une préoccupation croissante tant pour les producteurs que les utilisateurs de données de recherche. Les procédures d'authentification incorporées au processus de création, de transmission, de réception, d'utilisation, de maintien et de conservation des fichiers de données constituent la façon la plus efficace pour assurer l'authenticité des données au fil du temps. Actuellement, nous ne disposons ni de normes nationales de cette sorte ni d'aucun organisme pour surveiller leur application;
- Reformuler et articuler, à l'échelle nationale, des normes de sécurité pour protéger l'exactitude et l'uniformité des données. Ces normes devraient faire état : (1) de méthodes pour déterminer l'actif des données et de procédures de gestion du risque pour évaluer la vulnérabilité; (2) la détermination des exigences sur les plans juridique, législatif, de la réglementation et des obligations contractuelles et inclure des directives en matière d'éthique et des droits de propriété intellectuelle; (3) un ensemble de principes, de méthodes et de procédures que doivent suivre les organismes pour assurer la fiabilité en ce qui concerne la création, le maintien sûr, l'utilisation confidentielle et la conservation authentique de leurs données.

Si le Canada créait des archives nationales de données de recherche, seraient-elles utilisées? La vaste expérience acquise par d'autres pays démontre une croissance de l'utilisation des données, en fait de fréquence ainsi que de nombre d'utilisateurs (voir l'annexe C).

Parmi les dix ensembles de données les plus demandés par les utilisateurs des archives de données du Royaume-Uni, au cours de l'année financière 2000-2001, quatre des titres provenaient de ministères gouvernementaux, deux étaient subventionnés par des ministères gouvernementaux et quatre étaient parrainés par un important organisme subventionnaire de la recherche.

Rapport annuel du R.-U. sur les archives de données

Deux des éléments les plus utilisés pour mesurer les niveaux d'activité des archives de données sont l'envergure de leurs collections et le nombre de clients qu'elles desservent. L'an dernier, par exemple, le ICPSR de la University of Michigan a ajouté
1 835 fichiers de données à sa collection, soit une augmentation de 8 p. 100 par rapport à l'année précédente. D'autre part, le consortium a diffusé, en parallèle, 5 000 gigaoctets de données à ses clients. Durant la même période, les archives de données du R.-U. ont traité plus de 500 acquisitions et desservi 1 000 clients qui avaient passé 2 000 commandes, ce qui, au total, équivalait à près de 9 000 fichiers de données. Il s'agissait d'une augmentation de 2 000 fichiers de données livrés à des utilisateurs au cours d'une période de trois ans.

Plusieurs registres d'archives de données utilisent les statistiques basées sur le trafic Web. L'Oxford Text Archive, par exemple, a fait état de plus de 18 000 téléchargements de textes électroniques en 1999 et 2000. Cette utilisation de textes électroniques est de 39 fois supérieure aux commandes hors ligne de l'Oxford Text Archive. En plus du téléchargement de fichiers, le nombre de contacts des utilisateurs est aussi saisi à partir de statistiques en ligne. Par exemple, le ICPSR a signalé une croissance importante des contacts avec les clients en raison du plus grand nombre d'utilisateurs qui recouraient à Internet pour effectuer de la recherche et pour l'enseignement. Au cours des trois dernières années, alors que les ressources de ICPSR étaient disponibles en ligne, l'organisme a constaté une augmentation de l'accès aux données de plus de mille gigaoctets.

Les archives de données tiennent également des statistiques sur d'autres services. Par exemple, le programme de formation de l'ICPSR soutient, de façon constante, l'inscription annuelle de 500 à 540 participants. Pour citer un autre exemple, le Norwegian Social Science Data Service (NSD) tient des statistiques sur l'utilisation que font les chercheurs de leurs services afin de vérifier si les projets sont conformes à la loi. Le NSD déclare que ce service a connu une croissance allant jusqu'à 65 % dans une année donnée. Les services de référence tiennent habituellement leurs propres statistiques. En 2000-2001, les employés affectés aux archives de données du Royaume-Uni ont reçu 332 demandes de renseignements, suite aux commandes, relativement à des fichiers de données, ce qui ne représente qu'une facette des services de référence. Pendant l'année 2000, le service de données archéologiques a fait état d'un total de 174 demandes de renseignements qui incluaient des questions concernant l'utilisation du catalogue, l'aide technique, et autres renseignements généraux en archéologie. Le History Data Service reçoit environ 480 demandes de références générales pendant la même période. En plus d'offrir des services de référence, l'Oxford Text Archive a également fait l'évaluation technique de 125 demandes de subvention.

Une autre statistique utilisée par certaines archives de données est le volume des permis d'utilisation de logiciels élaborés et distribués par le biais de leurs services. Par exemple, des permis d'utilisation du logiciel NSDstat, mis au point et distribué par le NSD, ont été octroyés à environ

2 000 instituts en Norvège, et à 200 organismes à l'échelle mondiale. Bien que le nombre exact d'utilisateurs par permis soit inconnu, l'expérience démontre qu'un seul exemplaire de permis donne accès aux statistiques du NSD à plusieurs personnes.

Les archives de données plus importantes compilent aussi des statistiques sur leurs activités internationales. Par exemple, le personnel du German Central Archive for Empirical Social Research (ZA) déclare que chaque année, de façon continue, environ 50 chercheurs de calibre international mènent régulièrement, sur place, des recherches au moyen des données, à l'EUROLAB des ZA. Les ZA intègrent également les données et la documentation portant sur un certain nombre de projets internationaux, y compris le International Social Survey Program, pour le compte de 38 pays et sur les eurobaromètres, pour le compte de la Commission européenne.

Dans l'ensemble, l'expérience démontre que les archives de données qui offrent des services étendus (y compris la formation, l'élaboration de logiciels et l'accès en ligne aux fichiers de recherche) sont largement utilisées par les chercheurs d'envergure nationale et internationale, et dans chaque cas, on note une augmentation de l'utilisation.

Nombreuses sont les raisons sous-jacentes au partage des données résultant des études appuyées par le NIH. Le partage des données renforce les enquêtes scientifiques ouvertes, encourage la diversité des analyses et des opinions et les nouvelles recherches, facilite la vérification d'hypothèses et de méthodes d'analyse nouvelles ou de remplacement, appuie les études portant sur la collecte de données et les méthodes de collecte et d'analyse des données, encourage la formation de nouveaux chercheurs, facilite l'examen de thèmes non abordés par les chercheurs initiaux, et permet la création de nouveaux ensembles de données quand les données provenant de sources multiples sont combinées. En évitant de reproduire des activités coûteuses de collecte de données, le NIH est en mesure d'appuyer un plus grand nombre de chercheurs qu'il ne le pourrait si des données semblables devaient être recueillies à nouveau par chaque demandeur.

National Institutes of Health (États-Unis),
Énoncé de politique sur le partage des données de recherche

---

4. Les composantes de base des archives de données de recherche

Au cours des dernières années, le gouvernement du Canada a pris des mesures importantes en vue de créer une infrastructure de recherche globale et cohérente et un système de soutien de la recherche au Canada. Des mesures telles que la création de la Fondation canadienne pour l'innovation et de CA*Net3 ont réussi à combler plusieurs des lacunes existantes. Toutefois, l'un des besoins toujours à combler est celui d'une installation ou d'un institut qui serait responsable d'assurer la conservation des données de recherche et l'accès à celles-ci. Néanmoins, le Canada dispose déjà de nombreuses composantes de base pour la mise en place de cet organisme.

Services de traitement de données des universités – Les services existants de traitement de données des universités constituent peut-être la plus importante des composantes de base. Faute de ressources, ces services ne peuvent agir à titre d'organismes multiservices de traitement de données, mais ils possèdent le potentiel pour constituer des nœuds du réseau des archives nationales de données de recherche. Ce fait a été confirmé par l'expérience de l'Initiative de démocratisation des données, alors que les bibliothécaires et les archivistes de données de 66 universités se sont réunis pour former un consortium en vue d'améliorer l'accès aux données de Statistique Canada. Ces professionnels dynamiques gardent contact entre eux,, partagent les meilleures pratiques et l'information concernant les sources de données, les méthodes pour améliorer les services aux clients, de même que les progrès récents en matière de normes et de capacités techniques. Si les services de traitement de données des universités disposaient de ressources suffisantes, ils pourraient former un réseau national étendu de points de contact pour les chercheurs qui désirent accéder aux données de recherche compilées par d'autres, déposer les données qu'ils ont compilées eux-même, suivre des cours de formation afin d'acquérir des compétences supérieures en matière de traitement des statistiques et des données, et obtenir des conseils sur les meilleures pratiques et sur la façon d'agir conformément aux normes sur les données. Toutefois, le plus important serait qu'à long terme, le personnel responsable du réseau des services de traitement de données des universités puisse agir en tant que système de rétroaction pour les utilisateurs, aidant ainsi à structurer et à améliorer les services offerts par les Archives nationales de données de recherche, et, en bout de ligne, les connaissances créées par les chercheurs canadiens.

Instituts canadiens des archives – À l'instar des services de traitement de données d'universités, les instituts canadiens des archives, qui ont un mandat de recherche spécifique, offrent d'autres nœuds potentiels dans le cadre d'un réseau national d'archives de données de recherche. Ces instituts existent dans des environnements locaux, régionaux et institutionnels, soit comme entités indépendantes, comme partie d'une institution mère ou au sein des administrations gouvernementales, municipales, provinciales et fédérales. De plus, ils existent dans de nombreuses collectivités qui ne sont pas dotées d'universités. Bien que les archives canadiennes aient, jusqu'à tout récemment, traité principalement des dossiers non numériques, l'infrastructure de leur collectivité, les normes descriptives, les pratiques exemplaires, une expérience exhaustive en matière de protection de la confidentialité et du droit d'auteur, etc. constituent des assises solides qui leur permettront de développer des connaissances et du savoir-faire en vue de participer au réseau national de données de recherche.

Représentation à l'échelle internationale – Bien qu'ils n'aient pas d'autorité nationale, les employés qui s'occupent des services de traitement de données de certaines universités assurent actuellement l'une des principales liaisons du Canada avec de nombreux organismes internationaux responsables de la gestion des données de recherche et de l'élaboration de

normes internationales en vue de la création de métadonnées, du partage et de la conservation des données. La création de ces normes, les ententes et les pratiques communes sont essentielles dans un monde scientifique qui de plus en plus travaille hors frontières. L'expérience et les compétences de ces personnes peuvent être utilisées pour coordonner les efforts, protéger les intérêts du Canada et représenter ce dernier quand des décisions clés ayant des incidences à long terme sont prises.

Infrastructure de la transmission de données – CA*Net3 relie les services de traitement de données des universités et, sous peu, CA*Net4, le réseau optique national de transmission de données à très haute vitesse, développé par le Réseau canadien pour l'avancement de la recherche, de l'industrie et de l'enseignement (CANARIE) sera mis en service. CA*Net3, qui relie maintenant tous les instituts de recherche importants au Canada, fournit le pipeline nécessaire à la transmissions des données de recherche à l'échelle nationale. La capacité considérable de ce réseau permet la transmission rapide, efficace et fiable de très volumineux ensembles de données importants et complexes, ce qui est essentiel dans l'avenir. La dimension et la complexité des ensembles de données de recherche augmentent à un rythme effarant.

Cadres de gestion – Les cadres de gestion de l'utilisation des données de recherche sont tout aussi importants que les pipelines numériques et les nœuds d'accès. En raison des renseignements personnels délicats qu'elles contiennent, les données des sciences humaines, en particulier, doivent être gérées au moyen d'un plan d'ensemble et de cadres éthiques ainsi qu'en conformité avec la Loi sur l'accès à l'information et la protection des renseignements personnels. Les lignes directrices des trois conseils sur l'éthique de la recherche avec des êtres humains fournissent l'un de ces cadres. Ces directives font état, en termes généraux, des principes qui devraient guider les Archives nationales de recherche dans le traitement des renseignements personnels et de la confidentialité. Ainsi, avec les comités universitaires d'éthique de la recherche, nous possédons les règlements et la capacité institutionnelle pour assurer la protection des renseignements personnels propres à chaque citoyen. Ces comités déterminent les conditions en vertu desquelles des données sensibles peuvent être déposées et diffusées et ainsi incorporent une première étape de présélection aux fins des Archives nationales de données de recherche.

Recherche et développement – Dans notre univers numérique qui prend rapidement de l'essor, de nombreuses activités liées au maniement des données de recherche sont encore effectuées sans les connaissances suffisantes. À titre d'exemple, que l'on pense à l'assurance de la qualité, de l'authenticité et de la sécurité des données de recherche. Des Archives nationales de données seront en mesure de tirer parti des nouvelles connaissances résultant de la recherche de pointe dans ce domaine, y compris, par exemple, le projet InterPARES que finance le CRSH.

Instituts partenaires – Divers instituts peuvent jouer un rôle important dans le fonctionnement et les services des Archives nationales de données. Tant les Archives nationales du Canada que la Bibliothèque nationale du Canada ont, au fil des ans, développé des compétences appréciables par le truchement de leurs archives respectives ainsi que dans la transition de ces archives sous forme électronique. L'environnement du stockage, les normes descriptives, la migration des formats physiques et logiques et la protection du droit d'auteur ne sont que certains des secteurs où le partage des connaissances et la réalisation de projets conjoints seraient possible.

Données de recherche – Les données elles-mêmes constituent les composantes centrales d'un service de données de recherche. Il va de soi que tous les ensembles de données de recherche ne devraient pas nécessairement être conservés. Certains ensembles seront peu utiles après

l'achèvement d'un projet aux fins duquel les données ont été compilées; d'autres contiennent des indices personnels qu'il est impossible d'éliminer efficacement ou reproduisent simplement des données compilées ailleurs. Toutefois, la décision de conserver certains ensembles de données et d'en supprimer d'autres est une question préoccupante en archivistique et est cruciale pour assurer un partenariat efficace entre les chercheurs et les archivistes de données.

L'existence d'une grande quantité de données de recherche n'est pas remise en question. Dans la première phase des consultations, le groupe de travail a déterminé que les chercheurs subventionnés par le CRSH produisent annuellement, en moyenne, quelques 400 ensembles de données. Comme le CRSH peut appuyer une partie seulement de la communauté des chercheurs en sciences humaines, le nombre total des ensembles de données produits chaque année pourrait être de trois à quatre fois supérieur à ce nombre. Ne sont pas inclus les ensembles de données produits par des chercheurs en sciences naturelles, des spécialistes de la santé ou des ingénieurs de la recherche, mais il n'est pas déraisonnable d'estimer que de 4 000 à 5 000 ensembles de données sont produits annuellement, grâce aux fonds publics. Bien qu'il soit impossible de connaître des détails à ce sujet, ces ensembles représentent un investissement public de dizaines millions de dollars par année.

Données de recherche du gouvernement – Les recherches du groupe de travail concernant les archives d'autres pays ont révélé qu'en sciences humaines, le gouvernement produit des données de recherche qui sont souvent plus utilisées que les données produites par les chercheurs eux-mêmes. Un nouvel organisme pourrait jouer un rôle important en exploitant une installation de conservation des données, en offrant l'accès au catalogue de données de recherche compilées au gouvernement. Le groupe de travail a souvent entendu des témoignages indiquant que l'accès à ces informations est, dans l'hypothèse la plus optimiste, difficile, et exige beaucoup de temps, et dans l'hypothèse pessimiste, impossible. Toutefois, on estime que les ministères et organismes gouvernementaux tels que Statistique Canada, Développement des ressources humaines Canada (DRHC), Santé Canada, Ressources naturelles Canada, Environnement Canada, Justice Canada et de nombreux autres ministères dépensent plus d'un milliard de dollars annuellement pour la collecte des données. Une administration publique efficace permettrait de donner aux chercheurs des moyens efficaces et efficients d'utiliser ces données[2].

Services de conservation pour d'autres organismes de recherche – Les technologies actuelles de l'information facilitent considérablement notre accès aux renseignements numériques et notre capacité de les traiter et de les appliquer aux sujets de recherche qui sont d'une importance primordiale pour les Canadiens et les Canadiennes. Toutefois, la conservation à long terme des documents de recherche numériques est un secteur qui n'est pas resté à la hauteur tant du point de vue technologique qu'institutionnel. Actuellement, dans le milieu de la recherche, on met l'accent sur la compilation et l'accessibilité des données sur Internet . Les services de catalogage et de conservation des données interinstitutions sont souvent considérés comme étant d'importance secondaire

[2] Le programme canadien des changements à l'échelle du globe, le Panel sur les systèmes d'information et de données, L'analyse de données et de politiques et les obstacles à l'accès des données au Canada : Sujets de recherche sur le changement mondial, (Société royale du Canada, 1996), p. 7.

ou complètement ignorés. L'Institut canadien d'information sur la santé, le Centre canadien de la statistique juridique, le Système canadien d'information pour l'environnement, GéoConnexions, et la Stratégie de données communautaires du Conseil canadien de développement social, qui a été annoncée récemment sont tous dotés d'excellents systèmes d'accès aux données, mais n'ont pas une stratégie bien conçue et fermement appuyée de conservation des données à long terme. Un des rôles les plus importants que puissent jouer les Archives nationales de données de recherche est d'offrir des services de conservation de données et l'expertise pour ces services et de mettre en place de nombreuses autres initiatives d'accès aux données de recherche.

Dans le cas des recherches subventionnées par les fonds publics, on devrait exiger que les données produites, les outils de recherche utilisés, les cadres de recherche sur les concepts et l'échantillonnage etc., soient archivés et mis à la disposition des autres chercheurs. Cette mesure serait très importante dans la réalisation d'activités telles que l'incitation à la collaboration, les études longitudinales, les répliques d'études , les études comparatives, la création de concepts de questions « normatives » dans certains secteurs d'examen et d'analyses secondaires. En exigeant l'archivage et l'accessibilité des données , on encouragerait la transparence, l'obligation de rendre compte et la responsabilité. De plus, la prise en compte de telles données devrait devenir l'un des points principaux dans la planification de la « nouvelle » recherche de source directe – cessez de réinventer la roue et réalisez des travaux plus novateurs et créatifs.

Répondant au questionnaire

---

5. Vers un modèle institutionnel : leçons tirées de la scène internationale

Le groupe de travail a examiné toutes les archives de données de recherche existantes qui sont axées sur les sciences humaines (voir l'annexe C). Cet examen comprend des entrevues de face à face avec des directeurs d'organismes de traitement de données, l'analyse comparative des politiques et des règlements, l'examen des services, des mandats, des budgets et des structures de gestion. Les principales leçons tirées de l'examen sont les suivantes :

- Nombreux sont les pays qui ont reconnu depuis longtemps le besoin de se doter d'archives de données de recherche pour aider et soutenir les travaux de la communauté des chercheurs. Plusieurs archives de données examinées existent depuis 30 ans ou plus;
- Bien que de nombreux services d'archivage de données, en particulier les services d'accès et de formation, soient diffusés plus efficacement lorsque répartis dans un certain nombre d'installations, pour des raisons d'ordre économique, pratique, d'efficacité et de conservation, la gestion du réseau et des fonctions relatives à l'élaboration de normes est plus efficace lorsqu'elle est effectuée dans une seule installation;
- Les archives de données diffèrent l'une de l'autre. Chacune d'entre elles a été établie dans un contexte national ou disciplinaire spécifique qui reflétait les besoins particuliers du milieu scientifique qu'elle desservait. Il s'agit d'organismes intégrant des disciplines spécifiques, de petite à grande envergure, offrant des services restreints, jusqu'à des instituts nationaux pluridisciplinaires importants axés sur la R-D, qui offrent des services étendus et des réseaux à l'échelle internationale;
- Les archives de recherche florissantes sont directement reliées à l'infrastructure de recherche d'un pays, plutôt qu'au milieu des archives. Elles se caractérisent par l'orientation des services qui met en valeur l'accès aux données de recherche et la conservation des données les plus utiles aux fins de la recherche plutôt que sur la saisie de fichiers de données antérieurs;
- L'archivage des données de recherche est un travail complexe et hautement technique. Les archives de données exemplaires emploient des professionnels dévoués, spécialistes des données et accordent beaucoup d'importance à la formation des gestionnaires de la prochaine génération de données de recherche. La formation de personnel hautement qualifié répond aux besoins tant du milieu de la recherche qu'à ceux d'autres organismes des secteurs public et privé qui doivent traiter de grandes quantités de données;
- Il existe une corrélation directe entre le financement stable des archives de données de recherche et la qualité du soutien qu'elles offrent à la communauté des chercheurs. Par sa nature même, l'archivage est une entreprise à long terme. Les archives de données les plus utiles sont celles dont l'existence est assurée;
- Bien que l'archivage des données de recherche exige des engagements à long terme en matière de financement, les coûts institutionnels ne constituent qu'une très petite fraction des coûts de la collecte de données;
- L'établissement de liens de confiance avec les utilisateurs et les producteurs de données de recherche est essentiel. Si les utilisateurs ne peuvent se fier à la transmission efficace et en temps opportun de données de qualité et si les dépositaires ne sont pas convaincus que leurs droits de propriété intellectuelle seront sauvegardés et que la protection de leurs participants sera assurée, nul ne pourra faire confiance aux services fournis ou les utiliser;

- Les archives de données exemplaires sont souples et indépendantes sur le plan institutionnel. Elles travaillent en étroite collaboration avec de nombreux ministères gouvernementaux et avec des universités mais ne dépendent d'aucun organisme en particulier en matière de stabilité financière ou de prise de décisions. Les archives de données doivent être indépendantes afin d'assurer que les besoins de la communauté des chercheurs en matière d'accès aux données demeurent la principale priorité, plutôt que les besoins relatifs à la tenue des dossiers des ministères gouvernementaux ou des archives traditionnelles de données de recherche sur la culture. La souplesse est importante eu égard à l'adoption de nouvelles technologies et à la capacité de suivre l'évolution des besoins des chercheurs.

Le groupe de travail qui a fait un examen détaillé de 36 instituts a déterminé trois stratégies généralisées visant la conservation et l'accessibilité des données de recherche. Chaque stratégie possède les caractéristiques organisationnelles des services actuels offerts par les services nationaux d'archivage de données :

- **Des archives de données à petite échelle, spécialisées dans les données relatives à des sujets précis,** habituellement hébergées par un département universitaire qui possède une capacité limitée de maniement des données et qui utilise des technologies facilement disponibles sur le marché. La clientèle n'inclut parfois que des chercheurs d'une seule discipline ou d'un petit groupe de domaines de recherche et les budgets de fonctionnement annuels varient de 200 000 $ à 400 000 $.
- **Des archives de données de dimension moyenne, intégrées à un organisme,** dont l'organisme parrain est habituellement un institut national de recherche ou un ministère. Le mandat et les activités de collecte de données de ces archives, qui sont souvent situées sur un campus universitaire pour mieux desservir la clientèle de base composée de chercheurs, sont basés sur celles de l'organisme parent. Ces archives offrent une gamme de services relativement étendue et il arrive que les employés assument parfois des rôles de premier plan dans des organisations nationales ou internationales connexes. Les budgets annuels varient de 500 000 à 1,5 million de dollars.
- **Des archives de données de recherche étendues,** offrant des services à une grande variété de communautés, y compris des chercheurs en milieu universitaire, des analystes de politiques provenant d'organismes non gouvernementaux et gouvernementaux, des organismes publics de services archivistiques, et de simples citoyens. Ces archives de données qui sont souvent établies en vertu d'une loi sont reconnues à titre d'instituts nationaux responsables des principes généraux et des fonctions particulières décrits dans leurs actes fondateurs. Par le truchement d'un ou de plusieurs emplacements pratiques et de l'utilisation exhaustive d'Internet, une gamme étendue de services est offerte. Les services incluent souvent la formation spécialisée et générale, du soutien technique et de la R-D. Les capacités de gestion des données de ces archives sont considérables et souvent développées sur place. Ces organismes ont établi des relations de travail avec d'autres organismes nationaux et ministères, et le personnel est souvent composé de dirigeants d'associations internationales qui participent activement à des échanges internationaux de données. Les budgets annuels de ces organismes varient de 3 à 6 millions de dollars.

Avantages résultant du dépôt et de l'archivage des données :

* *Encouragent les enquêtes scientifiques ouvertes;*
* *Encouragent la diversité des analyses et des opinions;*
* *Favorisent les nouvelles recherches et permet la vérification des nouvelles méthodes ou des méthodes de remplacement;*
* *Améliorent les méthodes de collecte et d'évaluation des données en les soumettant à l'examen d'autres chercheurs;*
* *Réduisent les coûts en permettant d'éviter de reproduire les travaux de collecte de données;*
* *Constituent une ressource importante aux fins de la formation en recherche;*
* *Assurent la sauvegarde des données;*
* *Permettent aux propriétaires d'éviter les tâches administratives associées aux utilisateurs externes et à leurs requêtes;*
* *Remplissent les obligations voulant que lesrecherches subventionnées soient mises à la disposition de la communauté des chercheurs;*
* *Permettent aux chercheurs de démontrer l'utilisation continue des données après l'achèvement des travaux de recherche initiaux.*

Site Web du Inter-University Consortium
for Political and Social Research, University of Michigan

---

6. Principes fondamentaux et hypothèses

Le groupe de travail a conclu que des archives nationales de données de recherche devraient fonctionner en conformité avec un ensemble de principes fondamentaux. L'objectif global devrait être de créer « un système de confiance » offrant au milieu de la recherche des services accessibles et étendus qui permettent aux utilisateurs de localiser, de demander, d'extraire et d'utiliser les ressources en données de façon simple, transparente et économique. Un tel système devrait être conforme aux principes fondamentaux suivants :

1) Des archives nationales de données de recherche devraient encourager la création de connaissances en constituant une partie intégrale du processus de recherche et devraient également faciliter les découvertes et le processus décisionnel au Canada, y compris la formulation de politiques publiques, en conservant les sources des preuves et en facilitant l'accès à ceux-ci;
2) Dans la mesure du possible, l'accès aux données devrait être libre et gratuit;
3) L'assurance de la confidentialité et la protection de la vie privée et des êtres humains qui participent aux activités de recherche devraient être primordiales;
4) Le grand public devrait avoir accès aux données recueillies grâce aux fonds publics sous réserve des conditions relatives à leur utilisation préliminaire équitable par le dépositaire et des dispositions éthiques et juridiques régissant ladite collecte de données.

En maintes occasions, le groupe de travail a appris de sources qui font autorité que l'établissement d'un climat de confiance constitue l'élément clé de la création d'un système exemplaire de conservation et d'accessibilité de données de recherche. Ceci ne peut être accompli que si les utilisateurs et les dépositaires de l'organisme savent que les archives font partie intégrante de leur processus de recherche, qu'elles fourniront des services utiles et ajouteront de la valeur à leurs travaux. De plus, le service de traitement de données doit appuyer et se conformer activement aux règlements et directives établis concernant la protection de la confidentialité, de la vie privée et des droits de propriété intellectuelle. Dans la mise en place d'un système de confiance, les éléments les plus importants pour un nouvel organisme sont la stabilité à long terme, tant sur le plan de la structure institutionnelle que sur celui du financement. Il s'agit de leçons durement apprises par de nombreuses archives de données du monde entier. La source du mandat, de la régie, de l'obligation de rendre compte et un engagement à long terme en matière de ressources financières requises sont des éléments déterminants quant à la réussite ou à l'échec de l'entreprise.

7. Options pour le Canada

À la lumière de ces expériences et en consultation avec la communauté des chercheurs, le groupe de travail a conclu qu'il serait à l'avantage du Canada de créer un organisme possédant les caractéristiques générales suivantes :

- Mandat global établi en fonction des besoins d'une grande variété d'intervenants;
- La société et les personnes seraient l'objet central des données cibles;
- Service mettant en valeur la conservation et l'accessibilité des données ;
- Capacité d'assurer la protection de la vie privée et de la confidentialité comme élément fondamental de ses principes de fonctionnement;
- Capacité de traiter des données conformément aux normes internationales, participation aux échanges internationaux de données et représentation des intérêts du Canada lors de négociations internationales;
- Capacité d'effectuer de la recherche de pointe et de faire évoluer les sciences de l'information et de l'archivistique;
- Utilisation des techniques d'information et de communication les plus récentes en vue de maximiser l'accès aux données de recherche tout en réduisant les coûts en temps et en argent pour les chercheurs;
- Capacité d'éduquer et de former tant les producteurs que les utilisateurs des données de recherche et les gestionnaires de la prochaine génération de données;
- Établissement de relations de travail continues avec d'autres organismes nationaux, tels que les Archives nationales et la Bibliothèque nationale ainsi qu'avec des organismes comme CANARIE;
- Adhésions à des organismes et autres ententes officielles relatives aux échanges de données avec d'importantes archives de données à l'étranger, telles que l'ICPSR, aux États-Unis, et le réseau européen CESSDA;
- Financement public soutenu à long terme comme source principale de soutien. Ce financement pourrait être complété par la vente de produits à valeur ajoutée dérivés de données et de services de consultation à des organismes à but lucratif, mais ces ventes ne devraient pas constituer le financement de base.

Toutefois, dans le contexte canadien, des Archives nationales de données de recherche doivent également répondre aux critères suivants :

- Service entièrement bilingue;
- Accès aux données de recherche produites à tous les niveaux du gouvernement tout en respectant les limites de compétences fédérales et provinciales dans les secteurs comme l'éducation et la santé;
- Respect des lois, règlements et directives sur la propriété intellectuelle canadienne, du droit d'auteur, de la vie privée et de la confidentialité et collaboration à leur développement;
- Maintien de relations étroites de travail avec les principaux producteurs canadiens de données tels que Statistique Canada et les organismes de statistique provinciaux;
- Utilisation et soutien des infrastructures de recherche existantes, des services de soutien à la recherche, des programmes de soutien et de financement, y compris les services existants de traitement de données des universités et les bibliothèques de recherche, les conseils de soutien à la recherche, la Fondation canadienne pour l'innovation, le Centre canadien de la statistique juridique, l'Institut canadien d'information sur la santé ;

- Intérêt dans les données de recherche en sciences humaines et, le cas échéant, en sciences naturelles, en sciences de la santé et en génie.

L'archivage de données comprend un engagement à long terme en matière de ressources, d'expertise et des services publics requis afin d'assurer un accès continu aux fichiers de données, de décrire et de documenter des fichiers et d'offrir un accès à ces fichiers tout en assurant le contrôle intellectuel. L'une des raisons qui expliquerait pourquoi cette idée suscite peu d'enthousiasme chez les chercheurs est qu'il est difficile de savoir quelles données ont été recueillies. Ce serait tout à fait logique d'utiliser des économies d'échelle et de centraliser les ressources requises pour une entreprise de cette envergure.

épondant au questionnaire

Au Canada, les Archives nationales de données de recherche devraient répondre aux besoins en matière de conservation et d'accessibilité des données et repousser les frontières des sciences de l'information et de l'archivistique. Elles devraient avoir pour assise l'infrastructure de recherche existante tout en intégrant les leçons tirées de l'expérience en archivage de générations de données acquise dans d'autres pays. Il faudrait surtout que les archives soient bien adaptées au contexte social et institutionnel du Canada, tout en répondant aux exigences du public en ce qui concerne l'obligation de rendre compte et la régie efficace.

En étudiant les moyens à utiliser pour créer des Archives nationales de données de recherche, le groupe de travail a examiné les centres de recherche universitaires subventionnés par le gouvernement, a demandé l'avis du Bureau du Conseil privé et utilisé les directives figurant dans le Cadre d'examen des différents modes d'exécution de programmes publié par le Conseil du Trésor. Le groupe de travail a pris en compte six options possibles et a examiné chacune d'elles de façon exhaustive. Il a examiné les structures institutionnelles et de régie, les exigences liées au démarrage de l'organisation, à sa stabilité à long terme, ainsi que ses forces et ses points faibles selon la perspective des utilisateurs et des producteurs de données (voir l'annexe D).

En premier lieu, le groupe de travail a étudié l'option consistant à créer une nouvelle division ou un organisme de service spécial chapeauté par un établissement national existant, tel que les Archives nationales ou la Bibliothèque nationale du Canada. Bien que le mandat des Archives nationales soit assez étendu pour inclure les données de recherche non publiées, leur niveau actuel de financement ne leur permettrait pas d'intégrer un tel nouveau secteur de services, alors que l'organisme doit simultanément relever des défis importants, à l'échelle du gouvernement, en ce qui concerne la gestion de l'information dans les secteurs de l'administration électronique, de la transition de ses dossiers à une forme électronique et de la numérisation de ses ressources documentaires existantes. De plus, l'échec d'un essai antérieur en vue de créer une division d'archivage de données au sein des Archives nationales (1973-1986) laisse à penser qu'il existe une disjonction entre le rôle étendu de conservation de la culture qu'assument les Archives nationales et le rôle spécifique de prestation de services qu'assumeraient les Archives nationales de données de recherche, au sein de l'infrastructure de recherche du Canada. Ces différences tiennent des stratégies d'acquisition, des compétences du personnel, des programmes descriptifs courants, et des besoins de la clientèle.

La Bibliothèque nationale du Canada collectionne 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. Comme dans le cas des Archives nationales, l'accès aux ressources documentaires actuelles, leur conservation et leur maintien dans les deux organismes ne requièrent pas une connaissance approfondie des méthodes de recherche quantitative, des statistiques et de l'informatique, pour répondre aux besoins des personnes qui utiliseraient un réseau national de données de recherche. Le groupe de travail estime que les exigences particulières du milieu de la recherche relativement aux données requises pourraient marginaliser les activités des archives de données de recherche au sein de ces organismes et ainsi miner la stabilité à long terme requise pour assurer la réussite de l'entreprise.

Une autre option que le groupe de travail a examinée était la création d'un partenariat gouvernemental tel qu'évoqué par le Conseil du Trésor. Il s'agit de l'établissement d'un organisme comme partenariat entre les divers ordres des gouvernements fédéral et provinciaux. Bien que cette voie comporte certains aspects intéressants, elle ne se prête pas à l'établissement de liens directs avec la communauté des chercheurs des milieux universitaire et non gouvernementaux. Comme l'a appris le personnel des archives nationales dans d'autres pays, il s'agit d'un élément essentiel pour la création d'un système de confiance.

Un organisme distinct en vertu d'une loi ou un établissement public possède bon nombre des caractéristiques nécessaires à la création d'un système étendu, fiable et efficace d'archives nationales de données de recherche. Il s'agirait d'un organisme permanent protégé par la loi. Il constituerait un élément de la stratégie d'innovation, axé sur la recherche, le renforcement des capacités, la régie et la compétitivité internationale. Il aurait des directives précises en matière d'autorité et de responsabilité et serait parrainé par un ministère. Le financement serait assuré et stable et proviendrait d'une seule source. À l'instar de Statistique Canada, l'organisme aurait le potentiel et les moyens de se faire reconnaître comme étant un « système de confiance » et pourrait jouir du statut de représentant officiel du Canada sur la scène internationale. Le seul élément d'importance manquant est un lien direct avec les utilisateurs de données de recherche.

La dernière option étudiée, soit un centre situé dans une université, possède, sur place, ce lien direct et immédiat. Une telle installation donnerait un sentiment d'appartenance et le fonctionnement et les politiques seraient entre les mains des membres associés aux universités. L'organisme tire parti des services de données existants, l'expertise et l'infrastructure technologique en place dans les universités; il pourrait aussi utiliser un système hybride centralisé/décentralisé où le Centre s'occuperait de la conservation et du traitement des ensembles de données et où les membres associés offriraient des services locaux d'accès aux données, de dépôt de données, et agiraient à titre de conseillers sur place, d'organisateurs d'activités de formation; l'envergure de l'organisme serait mise à l'échelle et pourrait inclure les domaines scientifiques visés par le Conseil de recherches en sciences naturelles et en génie (CRSNG) et les Instituts de recherche en santé du Canada (IRSC). Enfin, des activités numériques de recherche en matière d'archives tireraient avantage de la proximité des chercheurs en sciences de l'information qui travaillent à l'université. La principale faiblesse de cette option est que l'organisme ne serait pas stable à long terme. Une deuxième faiblesse est le fait que l'organisme ne serait pas nécessairement habilité à agir à titre d'organisme national sur la scène internationale.

Après avoir examiné et débattu exhaustivement toutes ces options, le groupe de travail a conclu qu'en raison de la nature du système de gouvernement fédéral du Canada, des nouvelles technologies de communication et d'information, des caractéristiques particulières de la communauté des chercheurs et de l'essor des besoins de l'économie axée sur le savoir, une occasion unique d'innovation institutionnelle se présente – la création d'un organisme hybride combinant la stabilité d'un organisme distinct constitué par une loi et les liens avec les utilisateurs d'un centre de recherche situé dans une université.

8. Recommandations concernant la mise en place d'un réseau national d'archives de données de recherche

Afin de mettre en place un service national efficace d'archives de données de recherche qui répondrait le mieux aux besoins de l'économie canadienne axée sur le savoir, encouragerait l'innovation, utiliserait les capacités de l'infrastructure existante, assurerait une administration publique efficace et représenterait le Canada sur la scène internationale, le groupe de travail recommande que le gouvernement du Canada prenne les mesures suivantes :

- Crée, par voie législative, un réseau national d'archives de données de recherche en tant que version modifiée d'un organisme distinct constitué en vertu d'une loi;
- Exige que l'organisme relève du Parlement par l'intermédiaire soit du ministre de l'Industrie ou de la ministre du Patrimoine canadien, ou des deux, dans le cadre d'une structure de responsabilisation semblable à celle d'organismes tel le Secrétariat du changement climatique;
- Autorise cet organisme à fonctionner sans liens de dépendance, à l'instar des conseils de soutien de la recherche du gouvernement fédéral;
- Attribue directement des fonds d'exploitation à l'installation centrale et aux nœuds du réseau par des crédits parlementaires annuels dans le cadre du processus budgétaire ordinaire du gouvernement fédéral, ou encore alloue des fonds par l'intermédiaire des conseils fédéraux de soutien de la recherche participants, comme dans le cas des Réseaux des centres d'excellence.

En ce qui concerne les structures et les activités d'un réseau national d'archives de données de recherche, le groupe de travail fait, de plus, les recommandations suivantes :

- Que le nouvel organisme mette en place un réseau étendu de services et une installation centrale responsable de la gestion des données, de l'élaboration et de la conservation des normes et un ensemble de nœuds situés au sein des services de traitement des données de recherche des universités et autres établissements responsables de l'accessibilité et du dépôt des données et qui sont en mesure d'offrir des services de formation et de consultation aux chercheurs. On suggère que les établissements qui souhaitent devenir des nœuds forment un consortium en vue d'obtenir le financement initial requis auprès de la Fondation canadienne pour l'innovation et des fonds de contrepartie de divers organismes de recherche provinciaux pour l'élaboration de l'infrastructure;
- Que l'organisme crée un Conseil de gestion pour régir le réseau national d'archives de données de recherche. Le Conseil serait composé de représentants des diverses régions du Canada et de divers groupes intéressés qui gèrent, utilisent et produisent des données de recherche;
- Que l'organisme développe, au fil du temps, pour répondre aux besoins déterminés de la communauté des chercheurs, un système d'accès aux données de recherche et des services de gestion et de conservation des données;
- Que l'organisme développe une capacité à approfondir nos connaissances et notre compréhension des sciences de gestion de l'information, des cadres éthique et juridique, des pratiques de gestion de la technologie du savoir, et encourage une culture de partage des données , au sein de la communauté des chercheurs;

- Que l'organisme entretienne des relations formelles de travail collaboratif avec d'autres établissements nationaux tels que les Archives nationales et la Bibliothèque nationale du Canada et qu'il conclue des ententes relatives à la conservation et l'accessibilité des données de recherche avec des producteurs de données importants tels Statistique Canada et les organismes provinciaux de l'analyse statistique;
- Que l'organisme soit habilité à agir au nom du gouvernement du Canada lors de négociations internationales touchant les normes de gestion des données de recherche et des pratiques courantes.

Bien que le groupe de travail soit persuadé que le modèle décrit ci-dessus permettrait au Canada d'être à l'avant-garde dans le domaine de l'archivage de données et des sciences de l'information et qu'il augmenterait considérablement l'avantage concurrentiel du milieu de la recherche canadien, les membres du groupe sont également conscients que la solution idéale n'est pas toujours la plus pratique ou la plus réalisable et proposent deux autres solutions pour créer des archives nationales de données de recherche.

1) **Un réseau national d'archives de données de recherche sous la gouverne du CRSH** – Pour donner suite à la stratégie du Economic and Social Research Council, du Royaume-Uni, on créerait, sous la gouverne du CRSH, une installation et un réseau d'archives de données intégrés à une université. L'organisme relèverait du Conseil d'administration du CRSH et serait doté de la même gestion et structure en réseau, et d'une gamme de services semblables à ceux de l'option décrite précédemment. Une loi habilitante ne serait pas nécessaire dans le cas d'un tel organisme, car il serait intégré à la fonction de soutien à la recherche qui fait partie du mandat du CRSH, mais il n'aurait pas la stabilité à long terme que procurent les dispositions législatives. Il tirerait avantage de liens directs avec la communauté des chercheurs ainsi que des relations de travail que maintient le CRSH avec d'importants producteurs de données, tel Statistique Canada. On prévoit qu'il pourrait être mis à l'échelle afin d'inclure tous les secteurs de recherche scientifique et des sciences humaines et qu'il prendrait avantage de la recherche que finance le CRSH en sciences de l'information et en archivistique.

2) **Un organisme de service spécial chapeauté par les Archives nationales du Canada** – Bien qu'en apparence, cette voie puisse sembler la plus logique pour créer des Archives nationales de données de recherche, il est à noter que le groupe de travail n'a reçu que peu de suggestions recommandant l'adoption de cette voie. De plus, l'examen des archives de données de recherche dans d'autres pays a révélé que seules les archives de données de recherche danoises (Danish Data Archives)sont directement rattachées à des archives nationales et des renseignements empiriques laissent croire que cet arrangement a des effets défavorables. Néanmoins, la création d'un organisme de service spécial chapeauté par les Archives nationales serait une solution simple. Il pourrait tirer avantage de l'expérience archivistique du personnel des Archives nationales, utiliser les installations existantes ainsi que l'infrastructure technique et administrative et avoir le statut et l'autorité requis pour agir au nom du Canada dans le cadre du développement de normes et de pratiques internationales. En tant qu'organisme spécial, il jouirait d'une certaine autonomie au sein de la structure de gestion des Archives nationales tout en relevant de l'archiviste national du Canada. Cette voie donnerait une plus grande stabilité à l'organisme que celle de l'ancienne Division des archives ordinolingues. Le plus grand inconvénient de cette approche réside dans le fait que l'organisme n'aurait pas de lien direct et immédiat avec la communauté des

chercheurs, soit par le truchement de sa structure de gestion ou par les services de traitement de données des universités. Bien que ces deux éléments puissent être mis en place, l'organisme serait malgré tout intégré à un organisme fédéral dont la mission principale est de préserver la mémoire nationale et les dossiers du gouvernement et non de répondre aux besoins des chercheurs en matière de données.

Les chercheurs conviennent qu'une infrastructure permettant l'échange de données de recherche est depuis longtemps requise au Canada et que nous devons disposer d'une infrastructure cohérente pour la collecte, la documentation, le partage et la conservation des données numériques de recherche. En particulier, il est essentiel de réduire les coûts élevés de la collecte de données et d'assurer que les fichiers sont disponibles à des fins d'analyses secondaires.

Présentation de l'University of Calgary, Bureau du vice-président (Recherche)

---

9. Coûts d'un réseau national d'archives de données de recherche

Le niveau de financement requis pour l'établissement et le maintien d'un service national de données de recherche dépend de l'envergure de ses activités et de la gamme de services qu'il fournit. Le point crucial est de définir le niveau minimal de financement qui serait requis pour fournir des services adéquats. Un niveau trop faible, en plus de restreindre la gamme des services qui pourraient être offerts, menacerait également la reconduction du financement. Si le niveau de financement qui permet à l'organisme d'exister est trop bas, ce dernier risque d'être axé sur un nombre limité de disciplines ou de types de services et pourrait susciter du ressentiment de la part des utilisateurs qui ne recevraient pas les services voulus, et ainsi, la reconduction du financement serait compromise.

Le financement d'un réseau national d'archives de données de recherche – à la fois l'installation centrale et ses nœuds – devrait être assumé par le gouvernement du Canada. Bien que des fonds additionnels puissent être obtenus d'autres sources, telles les subventions à la R-D, la vente de produits à valeur ajoutée dérivés de données, et l'imposition de frais pour les services de consultation spécialisés, les activités générales de l'organisme devraient être financées par le gouvernement. C'est le seul moyen efficace d'assurer la stabilité à long terme des archives de données de recherche, qu'elles desservent tous les Canadiens et Canadiennes, dans toutes les régions, et qu'elles répondent aux besoins d'un large éventail de chercheurs et de producteurs de données de recherche en milieu universitaire, d'organismes gouvernementaux et non gouvernementaux et du secteur privé.

L'établissement précis des coûts d'un réseau national d'archives de données de recherche selon le modèle recommandé est impossible parce que le groupe de travail ne dispose pas des ressources requises. Toutefois, l'étude à l'échelle internationale des archives existantes a fourni des points repères importants quant aux niveaux de financement requis pour offrir certains niveaux de service.

En dollars canadiens courants, et une fois que l'organisme sera en service, le budget annuel de fonctionnement d'un organisme offrant des services étendus de traitement de données de recherche est d'environ 3 millions de dollars. Comme le souligne l'étude de faisabilité des archives de données de l'Irlande, environ 40 % des fonds seraient consacrés à l'acquisition, au traitement, au catalogage et à la conservation des données, tandis que les fonds restants (60 %) seraient dépensés pour le développement des procédés requis pour répondre aux besoins des utilisateurs[3].

Les coûts de l'infrastructure initiale dépendraient d'un certain nombre d'éléments, y compris l'emplacement et la dimension de l'installation centrale de conservation et de traitement des données, le nombre de nœuds du réseau, la répartition des fonctions spécifiques entre les nœuds et l'installation centrale, ainsi que la capacité et la complexité globales du matériel informatique. Si l'organisme était intégré aux instituts de recherche canadiens, plutôt qu'aux Archives nationales ou à un autre ministère du gouvernement fédéral, le financement de l'infrastructure pourrait provenir de la Fondation canadienne pour l'innovation et des fonds de contrepartie de divers organismes provinciaux.

[3] The Data Archive, University of Essex, "The Irish Data Archive Feasibility Project", 1997, p. 49.

Les services fournis par le réseau, et, par conséquent, ses coûts opérationnels, pourraient être augmentés au fil du temps, au fur et à mesure que croissent le nombre de dépôts et d'utilisateurs. C'est la voie habituelle qu'ont adoptée d'autres pays. Le volume de données conservées n'a pas d'incidences importantes sur les coûts de fonctionnement, car le prix du stockage numérique baisse rapidement. L'expérience des autres pays démontre plutôt que le maniement et la gestion des données, ainsi que les services à valeur ajoutée augmentent en fonction de l'utilisation des services par la communauté des chercheurs et du fait que ces derniers prennent conscience de leurs avantages réels et potentiels.

10. Prochaines étapes

Pour lancer le processus de création d'un réseau national d'archives de données de recherche, le groupe de travail fait les recommandations suivantes :

- Le Conseil d'administration du CRSH et l'archiviste national devraient créer un comité directeur pour la mise en œuvre du processus et demander le soutien de l'Association des universités et collèges du Canada (AUCC), de l'Association des bibliothèques de recherche du Canada (ABRC), de la Fédération canadienne des sciences humaines et sociales (FCSHS) et autres organismes connexes, former des parrainages avec des ministères si une loi habilitante est requise, et s'assurer de la participation des intervenants au développement plus poussé des mandats et structures organisationnelles;
- Ce comité devrait établir des contacts appropriés avec les hauts fonctionnaires du ministère de la Justice et du ministère des Finances et avec d'autres représentants de ministères et d'organismes pertinents afin de démarrer le processus de mise en place du nouvel organisme proposé et d'élaborer les détails de son fonctionnement;
- Ce comité directeur devrait être responsable de l'élaboration des critères de sélection pour l'installation centrale et des nœuds du réseau national des archives de données de recherche;
- Le comité devrait également offrir des conseils relatifs aux critères de sélection des membres du Conseil de gestion, au besoin;
- Le groupe de travail recommande fortement que les activités débutent dès que possible afin d'établir le bien-fondé du rôle que peut jouer un réseau national d'archives de données de recherche pour faire progresser davantage la stratégie d'innovation du gouvernement du Canada.

10.4 Déclaration sur l'accès aux données de la recherche financée par des fonds publics

Déclaration sur l'acces aux données de la recherche financée par des fonds publics

adoptée le 30 janvier 2004, à Paris
Les Gouvernements (1) de l'Allemagne, de l'Australie, de l'Autriche, de la Belgique, du Canada, de la Chine, de la Corée, du Danemark, de l'Espagne, des États-Unis, de la Fédération de Russie, de la Finlande, de la France, de la Grèce, de la Hongrie, de l'Irlande, de l'Islande, d'Israël, de l'Italie, du Japon, du Luxembourg, du Mexique, de la Norvège, de la Nouvelle-Zélande, des Pays-Bas, de la Pologne, du Portugal, de la République d'Afrique du sud, de la République slovaque, de la République tchèque, du Royaume-Uni, de la Suède, de la Suisse et de la Turquie.

Reconnaissant qu'un échange international optimum des données, informations et connaissances contribue de façon décisive aux progrès de la recherche scientifique et à l'innovation;

Reconnaissant que l'accès ouvert aux données et l'utilisation sans restriction de celles-ci favorisent le progrès scientifique et facilitent la formation des chercheurs;

Reconnaissant que l'accès ouvert aux données valorise les investissements publics portant sur la collecte de données;

Reconnaissant que l'accroissement substantiel des moyens de calcul permet l'exploitation de vastes quantités de données numériques financées sur fonds publics à de multiples fins de recherche par un grand nombre d'établissements de recherche du système scientifique mondial, ce qui accroît sensiblement le champ et l'ampleur de la recherche;

Reconnaissant les importantes retombées dont la science, l'économie et la société dans son ensemble pourraient bénéficier du fait des perspectives offertes par une utilisation accrue de données numériques, et le risque que des restrictions indues sur l'accès aux données de la recherche financée sur fonds publics et sur l'utilisation de ces données puissent diminuer la qualité et l'efficacité de la recherche scientifique et de l'innovation;

Reconnaissant qu'une disponibilité optimale des données de la recherche financée sur fonds publics pour les pays en développement renforcera la participation de ces pays au système scientifique mondial, et contribuera à leur développement économique et social;

Reconnaissant que la divulgation de données numériques issues de la recherche peut être restreinte par la législation intérieure sur la sécurité nationale, la protection de la vie privée des citoyens et la protection des droits de propriété intellectuelle et des secrets de fabrication, qui peut exiger des mesures de précaution additionnelles;

Reconnaissant que sur certains aspects de l'accessibilité des donnés financées par des fonds publics, des mesures additionnelles ont été prises ou seront introduites dans les pays Membres de l'OCDE et que des disparités dans les réglementations nationales pourraient entraver l'utilisation optimale des données financées sur fonds publics, aux échelons national et international.

Considérant les retombées bénéfiques de la mise en place des Lignes directrices de l'OCDE régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel (1980, 1985 et 1998) et des Lignes directrices de l'OCDE régissant la sécurité des systèmes et réseaux d'information (1992, 1997 et 2002) sur les politiques internationales à l'égard de l'accès aux données numériques;

DÉCLARENT S'ENGAGER A :
Œuvrer pour l'établissement de régimes d'accès aux données numériques de la recherche financée sur fonds publics conformes aux objectifs et principes suivants :

Ouverture : concilier l'intérêt d'un accès ouvert aux données permettant d'accroître la qualité et l'efficacité de la recherche et de l'innovation, et la nécessité de restreindre l'accès dans certains cas pour protéger des intérêts sociaux, scientifiques et commerciaux.

Transparence : rendre disponibles et accessibles au plan international les informations sur les organisations productrices de données, la documentation sur les données qu'elles produisent et les spécifications des conditions attachées à l'utilisation de ces données.

Légalité : prêter dûment attention, dans l'élaboration des régimes d'accès aux données numériques de la recherche, aux obligations légales nationales relatives à la sécurité nationale, au respect de la vie privée et à la protection des secrets de fabrication.

Responsabilité formelle : promouvoir des règles institutionnelles formelles explicites sur les responsabilités des diverses parties intervenant dans les activités liées aux données qui concernent la paternité, la mention des producteurs, la propriété, les restrictions concernant l'usage, les modalités financières, les règles éthiques, les conditions de licence et la responsabilité civile.

Professionnalisme : développer des règles institutionnelles pour la gestion des données numériques de la recherche basées sur les normes professionnelles applicables et sur les valeurs inscrites dans les codes de conduite des communautés scientifiques concernées.

Protection de la propriété intellectuelle : décrire les moyens d'obtenir un accès ouvert dans le cadre des différents régimes juridiques du copyright ou les autres législations sur la propriété intellectuelle applicables aux bases de données, ainsi qu'en ce qui concerne la protection des secrets de fabrication.

Interopérabilité : en coopération avec d'autres organisations internationales, prêter dûment attention à la nécessité de disposer de normes internationales pertinentes et à usage multiple.

Qualité et sécurité : décrire les bonnes pratiques quant aux méthodes, techniques et instruments employés pour le recueil, la distribution et l'archivage accessible des données afin de permettre un contrôle de qualité par un examen mutuel et d'autres moyens assurant l'authenticité, l'originalité, l'intégrité, la sécurité et la responsabilité.

Efficience : promouvoir une efficience accrue par rapport aux coûts au sein du système scientifique mondial par la description des bonnes pratiques relatives à la gestion de données et aux services spécialisés de soutien.

Responsabilité de rendre compte : évaluer le fonctionnement des régimes d'accès aux données, de manière à maximiser le soutien à l'accès ouvert pour la communauté scientifique et l'ensemble de la collectivité.

Rechercher la transparence dans les réglementations et politiques liées aux services d'information, d'informatique et de communications qui affectent la circulation internationale des données pour la recherche, et réduire les obstacles inutiles à l'échange international de ces données;

Prendre les mesures nécessaires pour renforcer les instruments existants et selon que de besoin créer dans le cadre du droit international et des législations nationales de nouveaux mécanismes et de nouvelles pratiques destinés à faciliter la collaboration internationale dans l'accès aux données numériques de la recherche;

Appuyer les initiatives de l'OCDE visant à promouvoir le développement et l'harmonisation des approches des Gouvernements souscrivant à la présente Déclaration qui sont destinées à maximiser l'accessibilité des données numériques de la recherche;

Prendre en considération les répercussions possibles pour les autres pays, notamment les pays en développement et les économies en transition, dans la recherche de solutions aux questions d'accès aux données numériques de la recherche.

INVITENT L'OCDE :
A définir un ensemble de lignes directrices de l'OCDE fondées sur un ensemble convenu de principes destiné à faciliter l'accès dans des conditions optimales d'efficacité par rapport aux coûts aux données numériques de la recherche, pour approbation par le Conseil de l'OCDE à un stade ultérieur.

(1) Y compris la Communauté européenne

11 Addenda

11.1 Sigles et acronymes

Sigle ou acronyme	Terme développé
ABRC	Association des bibliothèques de recherche du Canada
ADN	Acide déoxyribonucléique
AUCC	Association des universités et collèges du Canada
BAC	Bibliothèque et Archives Canada
BIND	Biomolecular Interaction Network Database
CCDA	Centre canadien de données en astronomie
CIUS	Conseil international des unions scientifiques
CNADRS	Consultation nationale sur l'accès aux données de recherche scientifique
CNC CODATA	Comité national canadien pour CODATA
CNRC	Conseil national de recherches Canada
CODATA	Comité pour les données scientifiques et technologiques du CIUS
CRSH	Conseil de recherches en sciences humaines du Canada
CRSNG	Conseil de recherches en sciences naturelles et en génie du Canada
DIRD	Dépenses intérieures brutes de R-D
DPOSS	Digital Salomar Observatory Sky Survey
É.-U.	États-Unis d'Amérique
FCI	Fondation canadienne pour l'innovation
fMRIDC	functional Magnetic Resonance Imaging Data Centre (États-Unis).
HDF-S	Hubble (Telescope) Deep Field – South
IC	Industrie Canada
ICIST	Institut canadien de l'information scientifique et technique
IRSC	Instituts de recherche en santé du Canada
ISO	Organisation internationale de normalisation
LPRPDE	Loi sur la protection des renseignements personnels et les documents électroniques
NIH	National Institutes for Health (États-Unis)
OCDE	Organisation de coopération et de développement économiques
OMPI	Organisation mondiale de la propriété intellectuelle
ONG	Organisation non gouvernementale
OP	Chercheur principal
OST	Office of Science and Technology (R.-U.)
PI / DPI	Propriété intellectuelle / droits de propriété intellectuelle
PIB	Produit intérieur brut
R.-U.	Royaume-Uni
SciDIF	Forum sur les données et l'information scientifique (forum de planification du CIUS)
SCIE	Systèmes canadiens d'information sur l'environnement
SMIB	Système mondial d'information sur la biodiversité
SSGOT	Système des systèmes globaux des observations de la Terre
TI	Technologie de l'information
UE	Union européenne

11.2 Ordre du jour de la consultation (sommaire fourni aux participants)

Le lundi 22 novembre – MATINÉE – Première séance

Heure	Activité	Description
7 h 30	**Petit déjeuner**	Réseautage officieux
8 h 30	*Première séance :* **Planter le décor** *Animateur David Strong*	Bienvenue et présentation – **David Strong**
8 h 50		« A Vision for Scientific Research Access » – Arthur Carty *Présentation par **Ian Wilson, BAC***
9 h 10		« The Power of Accessibility » – **Chuck Hasel**, *Génome Canada*
9 h 40		« Implications and Possibilities of the Ministerial Declaration on Access to Publicly Funded Research Data » **Marie Tobin**, *Industrie Canada*
10 h		Pause café et débat non structuré
10 h 30		Introduction aux récentes expériences pour donner accès aux données de la recherche scientifique – **Janet Halliwell**
10 h 40		Leçons apprises : **Paul Uhlir** *(perspective américaine)* **David Moorman** *(perspective canadienne en sciences sociales)* **Charlyn Black** *(perspective de l'IRSC)*
12 h	**Déjeuner**	Salutations à des organismes parrains **Patricia Kosseim**, *IRSC* **Eliot Phillipson**, *FCI*

Le lundi 22 novembre – DÉBUT D'APRÈS-MIDI – Séance 2

Heure	Activité	Description
13 h	Séance 2 : **Occasions** *Animateur David Strong*	Examen des occasions *Séance dirigée*
14 h 10		Pause café
14 h 30		Examen des occasions en plénière *Débat dirigé*

Le lundi 22 novembre – FIN D'APRÈS-MIDI – Séance 3

Heure	Activité	Purpose
15 h	Séance 3 : Défis Animateur David Strong	Détermination des défis *Séance dirigée*
15 h 45		Examen des défis en plénière *Débat dirigé*
16 h 15		Présentation des ébauches de notice d'impact – **David Strong**
16 h 30	**Évaluation de nos progrès**	Mesurer le soutien de la collectivité « Devoirs »
18 h 30	**Cocktail**	Réseautage officieux *Les participants doivent se retrouver au Musée canadien des civilisations*
19 h 15	**Souper**	Discours-cadre – **Claire Morris**, *présidente, AUCC*

Le mardi 23 novembre – DÉBUT DE MATINÉE – Séance 4

Heure	Activité	Description
7 h	**Petit déjeuner**	Réseautage officieux
8 h	*Séance 4 :* **Élaborer une vision pour l'accès aux données de la recherche** *Animateur David Strong*	Examen, prévisions, vue d'ensemble
8 h 15		Une exploration du futur *Séance dirigée*
9 h 15		Examen du futur en plénière *Débat dirigé*
9 h 45		Pause café
10 h		Facteurs inhibiteurs de l'accès – Chuck Humphrey

Le mardi 23 novembre – FIN DE MATINÉE – Séance 5

Heure	Activité	Description
10 h 30	*Séance 5 :* **Détermination des facteurs inhibiteurs clés** *Modérateur David Strong*	Détermination des facteurs inhibiteurs *Séance dirigée*
11 h 30		Examen des facteurs inhibiteurs en plénière *Débat dirigé*
12 h	**Déjeuner**	Salutations des organismes parrains **Steve Shugar,** *CRSNG* **Michael Raymont**, *CNRC*

Le mardi 23 novembre – APRÈS-MIDI – Séance 6

Heure	Activité	Description
13 h	*Séance 6* Élaboration de mesures *Modérateur David Strong*	Mesures pour combattre les facteurs inhibiteurs *Séance dirigée*
13 h 45		Examen des mesures proposées en plénière *Débat dirigé*
14 h 15		Résumé de l'atelier et description des prochaines étapes – **David Strong**
14 h 45	**Clôture**	Remarque de clôture – **David Strong**

11.3 Participants à la consultation

Nom		Organisation ou institution	Domaine d'expertise
Cindy	Bell	Génome Canada	Génétique
Vijay	Bhargava	Université de la Colombie-Britannique	Génie électrique
Charlyn	Black	Université de la Colombie-Britannique	Sciences de la santé
Sharon	Buehler	Université Memorial	Sciences de la santé
Sheila	Chapman	Instituts de recherche en santé du Canada	Sciences de la santé
David	Crane	Journaliste indépendant	Journalisme
Josef	Cihlar	Centre canadien de télédétection	Environnement
Mark	de Jong	Centre canadien du rayonnement synchrotron	Photonique
Bernard	Dumouchel	Institut canadien de l'information scientifique	Information et technique
Kenneth	Edgecombe	Université Queen's	Administration universitaire
Carole	Estabrooks	Université de l'Alberta	Médecine
Alan	Evans	Institut neurologique de Montréal	Neurosciences
Louis	Fortier	Université Laval	Biologie
Dan	Gale	Canadian Microelectronics Corp.	Électronique
Thomas	Goldthorpe	University Heath Network (U de T)	Informatique médicale
Elizabeth	Griffin	CNRC – Institut Herzberg d'astrophysique	Astrophysique
Chuck	Hasel	Génome Canada	Génétique
Vicky	Hipkin	Université de Toronto	Astrophysique
Christopher	Hogue	Hôpital Mount Sinai	Génétique
Gregory	Kealey	Université du Nouveau-Brunswick	Administration universitaire
Richard	Keeler	Université de Victoria	Administration universitaire
Martin	Kryzwinski	Genome Sciences Centre	Génomique
Ian	Lancashire	Université de Toronto	Sciences sociales
Andreas	Laupacis	Institut de recherche en services de santé	Médecine
Marc	Lepage	Génome Canada	Génétique
Bryan	Lynch	Université St. Francis Xavier	Chimie
Susan	McDaniel	Université de Windsor	Administration universitaire
Paul	Melancon	Université de l'Alberta	Biologie
Ikechi	Mgbeoji	Université York	Droit – PI
Javad	Mostaghimi	Université de Toronto	Administration universitaire
Francis	Ouellette	Bioinformatics Centre de l'Université de Colombie-Britannique	Génétique
David	Phipps	Université York	Administration universitaire
Linda	Pilarski	Université de l'Alberta	Sciences de la vie
Robert	Prince	Université York	Physique
Jorg	Sack	Université Carleton	Informatique
Dennis	Salahub	Université de Calgary	Administration universitaire
David	Schade	CNRC – Institut Herzberg d'astrophysique	Astrophysique
Christoph	Sensen	Université de Calgary	Génétique
Frances	Sharom	Université de Guelph	Sciences de la vie
Pamela	Slaughter	Institut de recherche en services de santé	Science de la santé
Randall	Sobie	Université de Victoria	Physique
Elizabeth	Spangler	Université de l'Île-du-Prince-Édouard	Médecine vétérinaire
Dale	Swan	Environnement Canada	Environnement
Fraser	Taylor	Université Carleton	Administration universitaire
Lucy	Thompson	Université du Nouveau-Brunswick	Sciences spatiales

Kip	Tyler	Manitoba Land Initiative	Géographie
Robert	Vet	Environnement Canada	Météorologie
Donald	Weaver	Université Dalhousie	Environnement
Alan	Wildeman	Université de Guelph	Administration universitaire
Michael	Wolfson	Statistique Canada	Sciences de la santé
Tsoi	Yip	Environnement Canada	Météorologie

11.4 Membres du Groupe de travail

David Strong
Président du Groupe de travail de la CNADRS
Président, Université Canada West

John ApSimon
Conseiller spécial auprès du président
Université Carleton

Guy Baillargeon
Agriculture et Agroalimentaire Canada

Joyce Garnett
Université Western Ontario
Présidente de l'Association des
bibliothèques de recherche du Canada

Charles (Chuck) Humphrey
Université de l'Alberta et
représentant de l'Université de l'Alberta
auprès de l'Association des bibliothèques
de recherche du Canada

Steven Jones
Centre de la recherche sur le cancer de C.-B.

Ellsworth LeDrew
Université de Waterloo

Peter Pennefather
Université de Toronto

Andrew Pollard
Université Queen's

Richard Rachubinski
Université de l'Alberta

John Rodgers
Toth Information Systems

John Spray
Université du Nouveau-Brunswick

Robyn Tamblyn
Université McGill

Paul Uhlir
US National Academies

11.5 Membres du groupe de gestion de projet

Gordon Wood
Président du Groupe de gestion de projet
Conseil national de recherches Canada

Jac van Beek
Conseil national de recherches Canada

Catherine Betz
Conseil national de recherches Canada

Dennis Blinn
Recherches en sciences et en génie Canada (CRSNG)

Tony Damiani
Industrie Canada

Stephanie Delorme
Conseil national de recherches Canada

Peter Leach
Leach Technologies Ltd.

David Strong
Président du Groupe de travail de la CNADRS
Président, Université Canada West

Alan Le Couteur
Recherches en sciences et en génie Canada (CRSNG)

David Moorman
Conseil de recherches en sciences humaines du Canada

Katy Nau
Fondation canadienne pour l'innovation

Glen Newton
Conseil national de recherches Canada

Stephanie Robertson
Instituts de recherche en santé du Canada

Alexandra Talbot
Conseil national de recherches Canada

ISBN 0-662-79595-4

9 780662 795957

IRSC CIHR
Instituts de recherche en santé du Canada
Canadian Institutes of Health Research

CRSNG
NSERC

Fondation canadienne pour l'innovation
Canada Foundation for Innovation

CNRC-NRC