



Canadian Association  
of Research Libraries

Association des bibliothèques  
de recherche du Canada

## Comblent les lacunes de données de recherche Examen de nouveaux services de bibliothèques

Mars 2010

Par Kathleen Shearer et Diego Argáez

Association des bibliothèques de recherche du Canada (ABRC)

Au nom du Sous-comité sur la gestion des données de  
l'ABRC :

Marnie Swanson (présidente, University of Victoria)

Pam Bjornson (CNR-ICIST)

Lynn Copeland (Simon Fraser University)

Michelle Edwards (University of Guelph)

Bernie Gloyn (observateur, Statistique Canada)

Margaret Haines (Carleton University)

Janine Schmidt (McGill University)

Kathleen Shearer (attachée de recherche, ABRC)

Diego Argáez (secrétaire, ABRC)



© 2010 ABRC, Le contenu dans ce document est sujet à une licence de Creative Commons Paternité - Pas d'Utilisation Commerciale 2.5 Canada.

**Table des matières**

**Introduction**

**Sensibilisation et promotion**

**Soutien et formation**

**Accès et découverte**

**Archivage et conservation**

**Environnements de recherche virtuelle**

**La voie à suivre**

**Références**

## **Introduction**

Comme l'affirmait John Wilbanks, vice-président de Science Commons, « nos capacités de mesurer, stocker, analyser et visionner les données est de plus en plus la nouvelle réalité à laquelle la recherche devra s'adapter<sup>1</sup> ». C'est vrai non seulement pour les sciences, les données de recherche dans tous les domaines sont de plus en plus reconnues comme actifs qui ont une valeur qui va au-delà des objets pour lesquels elles ont été créées à l'origine.

La plus grande partie des données auront été créées numériquement et elles seront stockées et gérées sur support électronique, de sorte qu'il sera facile de les rendre accessibles, de les reproduire et de les combiner avec d'autres données. En 2006, à la suite d'une consultation nationale sur les données de recherche au Canada, on envisageait « un nouveau monde de recherche et un monde tout à fait nouveau<sup>2</sup> ». Le rapport connexe poursuit en affirmant que les bases de données sont regroupées à l'intérieur des disciplines et des pays et entre ceux-ci, les connaissances peuvent faire de très grands bonds en avant qui transforment notre compréhension de la vie, du monde et de l'univers. Afin de favoriser ce « nouveau monde », les données doivent être créées et conservées d'une manière qui répond bien aux objectifs de la conservation à long terme. Pour ce faire, il faut assurer une gestion active pendant tout le cycle de vie des données, à partir du moment où elles sont envisagées pour la première fois. Il existe de plus de très grandes variations dans les habitudes d'utilisation des types et des formats de données de recherche dans l'ensemble des disciplines, ce qui ajoute à leur complexité.

En cette époque de communications savantes axées de plus en plus sur les données, les rôles et responsabilités n'ont pas encore été bien établis et les bibliothèques ont des occasions de mettre au point des services uniques et précieux. À bien des égards, les bibliothèques sont bien placées pour appuyer l'égérance des données de recherche. Dans les universités, elles sont déjà reconnues comme agents conservateurs et fournisseurs d'accès à d'autres types de contenus et elles ont de solides liens avec l'ensemble des disciplines. Toutefois, les données soulèvent également des difficultés spéciales. Les données de recherche présentent des formats et une utilisation très différents, il est donc impossible de les traiter comme type unique de document. De plus, les « collections de données »<sup>3</sup> qui peuvent être utilisées au-delà du projet de recherche original pour lequel elles ont été créées deviennent des ressources communautaires. Ainsi, il peut être difficile de démontrer que les services de données de bibliothèque profitent directement aux chercheurs d'un établissement quelconque.

Le présent document donne les résultats d'un examen du potentiel insoupçonné qui s'ouvrent aux bibliothèques dans le domaine des services de données de recherche. Les activités ont été cernées au moyen d'une analyse documentaire et du dépouillement des projets réalisés dans les bibliothèques et dans d'autres établissements du monde entier. Pour les besoins du présent rapport, les services de données de recherche ont été organisés en cinq secteurs distincts (il faut toutefois noter qu'il existe d'importants chevauchements entre ces secteurs) : sensibilisation et promotion, soutien et formation, accès et découverte, archivage et conservation et environnements de recherche virtuelle. Chaque secteur comporte une description générale accompagnée d'un certain nombre d'exemples. Ces exemples ne sont pas censés être un compte rendu exhaustif des projets actuels mais plutôt servent à mettre en lumière la gamme des possibilités.

---

<sup>1</sup> Wilbanks, J. p. 209

<sup>2</sup> Strong, et al. p. 1

<sup>3</sup> La National Science Foundation définit les collections de données de référence ainsi : « elles visent à servir de vastes segments des milieux scientifiques et de l'éducation. Parmi les caractéristiques de cette catégorie de collections numériques, mentionnons un vaste éventail et un ensemble divers de groupes d'utilisateurs, c'est-à-dire des scientifiques, des étudiants et des éducateurs provenant de disciplines, d'établissements et de lieux très différents. ». Consulter : National Science Foundation, annexe D.

### **Sensibilisation et promotion**

Les nombreuses consultations et études qui ont eu lieu au cours de la dernière décennie portaient sur la nécessité d'une meilleure conscientisation face à la gérance des données avec diverses collectivités d'intervenants. Les bibliothèques sont en bonne position pour assumer ce rôle, surtout celui de promouvoir la conscientisation auprès des chercheurs et des administrateurs d'université.

Les bibliothèques peuvent entreprendre des activités de conscientisation afin de promouvoir l'gérance des données. Parmi ces activités, mentionnons les conversations avec les intervenants et les chercheurs sur le campus, de distribution de documents d'information, l'organisation de conférences ayant pour thème l'gérance des données et la nomination de certains membres du personnel de l'université comme champions à l'intérieur des départements. Compte tenu du fait que les questions et les difficultés de la gérance des données peuvent varier grandement selon la discipline, les programmes de conscientisation sont les plus efficaces lorsqu'ils sont adaptés à une collectivité précise. Les messages doivent également être clairs et uniformes et il faudrait prendre en compte les questions culturelles des différentes disciplines qui pourraient devenir des obstacles à l'gérance généralisée des données de recherche.

### **Exemples :**

#### **Points de discussion en matière de science électronique à l'intention des doyens et directeurs de l'ARL**

[www.arl.org/bm~doc/e-science-talking-points.pdf](http://www.arl.org/bm~doc/e-science-talking-points.pdf)

Ce document a été élaboré pour les doyens et directeurs de la Association of Research Libraries (ARL) afin de les aider à mieux comprendre la science électronique et la cyberinfrastructure. Ce document définit les termes essentiels et décrit certaines des difficultés que soulèvent les données. Il y est également question de politiques pertinentes aux données (surtout aux États-Unis) et des rôles des bibliothèques à faire évoluer les environnements de science électronique.

#### **Trousse de sensibilisation à la gestion des données de l'ABRC**

[www.carl-abrc.ca/new/pdf/data\\_toolkit\\_communique\\_jan2010.pdf](http://www.carl-abrc.ca/new/pdf/data_toolkit_communique_jan2010.pdf)

La trousse de sensibilisation *Les données de recherche: un potentiel insoupçonné* a été mise au point pour permettre aux directeurs de bibliothèques de faire connaître les difficultés de la gestion des données aux administrateurs et aux chercheurs universitaires. Les données sont des actifs précieux qui, dans certains cas, ont un potentiel de réutilisation illimité. La trousse de sensibilisation met en évidence la nécessité d'assurer que les données de recherche soient gérées pendant tout leur cycle de vie pour qu'elles demeurent compréhensibles et utilisables. *Les données de recherche: un potentiel insoupçonné* donne aux lecteurs une compréhension générale de l'état actuel des données de recherche au Canada et ailleurs dans le monde. Ce document se divise en sept parties : Vue d'ensemble, Principaux avantages de la gestion des données, Contexte actuel, Études de cas, Lacunes dans la gérance des données au Canada, Politiques de gestion des données, Réponses aux questions que se pose le corps professoral et les administrateurs, Qu'est-ce qui peut être fait sur le campus?

#### **Vérification des données à Oxford University**

[www.ict.ox.ac.uk/odit/projects/digitalrepository/](http://www.ict.ox.ac.uk/odit/projects/digitalrepository/)

La vérification des données permet également d'améliorer la sensibilisation face aux difficultés des chercheurs. Les vérificateurs recueillent l'information au sujet des actifs de données actuels : les endroits où ils se trouvent, comment ils sont gérés, lesquels doivent être conservés à long terme et si oui ou non les pratiques actuelles mettent les données en danger. Un cadre de vérification des données a été mis au point par l'Union européenne et utilisé par la bibliothèque d'Oxford University (et d'autres) afin d'en connaître davantage sur les pratiques des chercheurs concernant les données et saisir leurs besoins d'utilisation en matière de services de soutien de leur gestion des données.

### **Diverses publications**

Il existe un certain nombre d'articles publiés récemment qui font la promotion de la gérance et du partage des données. Ces articles se rapportent directement aux préoccupations des chercheurs et constituent des outils précieux qui visent à renseigner le corps professoral au sujet de l'importance du partage et de la gérance des données. Voici quelques-unes de ces publications :

1. Special section on Data Sharing. *Nature*, v. 461, no. 7261, 10 September 2009
2. *Data publication : towards a database of everything*. Vincent S Smith; BMC Research Notes v. 2, <http://www.biomedcentral.com/1756-0500/2/113>
3. *Data Sharing : The Next Generation*. Declan Butler; *Nature* v. 446, p. 10-11, 1<sup>er</sup> mars 2007 : [www.nature.com/nature/journal/v446/n7131/full/446010b.html](http://www.nature.com/nature/journal/v446/n7131/full/446010b.html)

### **Formation et soutien**

En plus des activités de promotion et de sensibilisation, il existe un réel besoin de formation et de perfectionnement des compétences dans la collectivité de recherche. Les chercheurs ont besoin de conseils au sujet des plans de gestion de données, des normes techniques, du catalogage des données, des normes et processus concernant les métadonnées et la gestion de la conservation. En 2008, le Groupe de travail canadien sur la stratégie des données de recherche a publié une analyse des lacunes qui concluait que les chercheurs ont très peu des compétences nécessaires pour bien gérer leurs données. La situation est semblable dans les autres pays. Par exemple, un sondage effectué récemment au Royaume-Uni a conclu que « bien qu'il y ait des écarts entre les disciplines, la compétence des chercheurs en matière de gestion des données est en général mauvaise »<sup>4</sup>.

Actuellement, il y a peu de possibilités pour les chercheurs au Canada d'obtenir la formation en gestion des données. Le Groupe de travail sur la stratégie des données de recherche dirigé par l'ICIST (Institut canadien de l'information scientifique et technique) élabore en ce moment un cours sur les données de recherche et il existe actuellement un certain nombre d'occasions propres à chaque discipline pour donner une formation au sujet des données. Les bibliothèques, qui ont déjà signalé que les chercheurs les ont approché pour obtenir des conseils et de l'aide, pourraient combler les lacunes en permettant l'accès à des ressources de formation. Diverses stratégies peuvent être utilisées pour améliorer le niveau de compétence des chercheurs, y compris la rédaction de guides et la mise en place de pages Web « explicatifs », l'élaboration de tutoriels de formation et la mise en place de cours sur la gestion des données. Il faut noter que, en raison de la nature diversifiée des données de recherche, il faudra élaborer des ressources de formation de plus haut niveau orientées vers chaque discipline.

### **Exemples :**

#### **Bibliothèques du MIT : Gérer vos données**

<http://libraries.mit.edu/guides/subjects/data-management/index.html>

Le MIT a établi un site Web très complet qui fournit aux chercheurs des conseils sur la gestion de leurs données de recherche. La page d'accueil indique ce qui suit « Il est essentiel de gérer vos données avant de commencer votre recherche et pendant toute leur durée de vie afin d'assurer qu'elles sont toujours utilisables et d'être en mesure de les conserver et d'y accéder à long terme. » Le site comporte également une « liste de contrôle de la gestion des données » et offre des conseils pour mettre au point des plans de gestion des données. Les autres sections du site Web sont intitulées : Why Manage Your Data?, What is Data?, Evaluate Your Data Needs, Documentation and Metadata, File Formats, Organizing Your Files, Backups and Security, Sharing Your Data, Citing Data, Data Integration, Funding Requirements, Ethical and Legal Issues.

---

<sup>4</sup> Lyon, L. p. 54

### **Guide national de service des données de l’Australie**

<http://ands.org.au/guides/index.html>

Ce guide fournit les renseignements au sujet de certaines questions fondamentales sur la recherche axée sur les données et la gestion des données de recherche. Il s’agit de lignes directrices et de listes de contrôle qui fournissent et élargissent la gamme des possibilités pour les chercheurs, les gestionnaires de données et les organismes de recherche. Les guides qui sont actuellement disponibles par l’ANDS sont les suivants: Copyright and data, Creative commons and data, Data citation, Data management planning, File formats, Identify My Data, Metadata, Persistent identifiers, Register My Data, Research Data Australia, Research data policy and the Australian Code for the Responsible Conduct of Research, Scholarly communications, and Storage.

### **Forum sur la gestion des données de recherche : Digital Curation Centre**

[www.dcc.ac.uk/events/data-forum-2010-march/](http://www.dcc.ac.uk/events/data-forum-2010-march/)

Les forums sur la gestion des données de recherche ont été mis en œuvre face aux difficultés que posait le nouvel environnement numérique en rapide évolution. Ces forums ont pour principe sous-jacent de promouvoir l’échange d’expérience et de compétences en matière de gestion et de conservation des données de recherche comme moyen d’en améliorer la qualité, la fiabilité, le traitement, la gestion et l’accès. À ce jour, cinq forums ont été organisés depuis 2007. Chacun permet de choisir une question stratégique afin d’établir un cadre pour ses activités tout en demeurant ancré dans les préoccupations pratiques de la prestation des services de gérance des données.

### **IGDM02 – Cours sur la gestion des données à l’Australian National University**

<http://ilp.anu.edu.au/dm/>

Ce cours s’appuie sur le Guide de gestion des données (GD) élaboré par l’Australian National University (ANU). Il présente une introduction à la gestion des données. Les participants apprennent la nature des données de recherche numérique et pourquoi il faut les gérer. On discute des avantages de la gestion des données de même que les politiques des organismes de financement qui se rapportent aux données. Un certain nombre d’idées concernant la GD font l’objet de discussions détaillées, comme la sauvegarde, le contrôle de la version, les métadonnées et l’archivage. Pendant le cours, les participants doivent également élaborer un plan provisoire de gestion des données pour leur actuel travail de recherche. De plus, on explique les problèmes qui se rapportent aux données et les diverses méthodes de les régler à l’aide d’outils logiciels et les services de l’ANU tels que décrits.

### **Gérer et partager les données : Guide sur les pratiques exemplaires à l’intention des chercheurs.**

[www.data-archive.ac.uk/news/publications/managingsharing.pdf](http://www.data-archive.ac.uk/news/publications/managingsharing.pdf)

Publié par UK Data Archive, ce guide est conçu pour aider les chercheurs et les gestionnaires de données de toutes les disciplines et environnements de recherche à faire en sorte que leurs données sont de la meilleure qualité possible et ont le meilleur potentiel de réutilisation à long terme. Dans l’avant-propos on note qu’une bonne gestion des données est le fondement d’une bonne recherche. Si les données sont bien organisées et conservées, et que leur exactitude et leur intégrité sont contrôlées en tout temps, il en résulte des données de haute qualité, une recherche efficace et des économies de temps et de ressources. Les chercheurs eux-mêmes bénéficient grandement de la bonne gestion de leurs données de recherche. La gestion des données doit être planifiée dès le début de la recherche. Si la gestion devient partie intégrante des méthodes de recherche standard, elle n’encourt pas du temps ou des coûts supplémentaires.

### **Découverte et accès**

Bien que les données de recherche ont une valeur qui va souvent au-delà de l’objet pour lequel elles ont été créées, la plupart ne sont toujours ni partagées ni disponibles au public. Selon l’analyse des écarts de 2008, « Une bonne partie des données de recherche produites aujourd’hui sont difficiles d’accès par les

## Comblent les lacunes de données de recherche : Examen de nouveaux services de bibliothèques

autres milieux de recherche canadienne et ne sont pas souvent idéalement structurées pour être aussi utiles ou ouvertes que possible, même à l'intérieur de la discipline pour laquelle elles sont réalisées »<sup>5</sup>.

Les métadonnées sont particulièrement importantes pour faire en sorte que les données sont comprises par d'autres et sont réutilisées. Selon le National Science Foundation, « Pour que les données soient utiles, il est nécessaire de conserver une documentation adéquate qui se rapporte au contenu, à la structure, au contexte et à la source (c.-à-d. paramètres expérimentaux et conditions environnementales) de la collection des données – que l'on nomme collectivement *métadonnées*. Idéalement, les métadonnées constituent un enregistrement de tout ce qui peut intéresser un autre chercheur »<sup>6</sup>.

Les bibliothèques pourraient jouer un rôle essentiel dans l'élaboration de services visant à améliorer l'accès aux ensembles de données produites sur les campus ou ailleurs. Parmi ces services, on pourrait inclure le catalogage et des liens à divers ensembles de données conservés ailleurs. Ou, plus généralement, le service pourrait comprendre toute une gamme d'activités de conservation des données.

### Exemples :

#### **Enregistrer mes données auprès du Australian National Data Service**

<http://www.ands.org.au/services/register-my-data.html>

Le service Register My Data de l'ANDS permet aux chercheurs d'enregistrer des collections de données de recherche. Les descriptions des collections enregistrées sont publiées dans un certain nombre de services de découverte et d'application de recherche en ligne. Register My Data aide les chercheurs et les organismes de recherche à publier leurs collections de données de recherche, permettant la vérification des conclusions de recherche et la réutilisation des données de recherche précieuses. L'ANDS vise à contribuer à la mise en commun mondiale des données de recherche.

Le service Register My Data de l'ANDS est censé être un registre complet de la recherche effectuée en Australie. On y enregistre la description des collections de données vise à faire en sorte que les collections enregistrées s'affichent bien en vue dans les résultats des moteurs de recherche. Afin de maintenir la modernité des renseignements, ce service appuie un certain nombre de protocoles dynamiques d'échange et de cueillette visant à automatiser les communications avec les environnements de données gérées. Ce service appuie également des descriptions et les enregistrements manuels.

#### **Centre de conservation des données réparties (D2C2) dans les bibliothèques de Purdue University**

<http://d2c2.lib.purdue.edu/>

Le Digital Data Curation Center (D2C2) a pour objet de régler les problèmes de conservation et d'attaquer les problèmes qui se rapportent aux données, au flux de données et aux environnements de données non organisés, disparates, hétérogènes et répartis. Cet organisme collabore étroitement avec d'autres organismes, centres et groupes qui font des travaux de même nature afin que les pratiques et les normes puissent être partagées, examinées et évaluées. La plupart des projets portent sur les partenariats avec des scientifiques de certains domaines et des technologues de l'information dans le but de répondre aux besoins de données de la collectivité de recherche dans le vrai monde. La portée des projets s'étend en général sur différents secteurs y compris la gestion des données, la découverte et la diffusion des données, la description des données et leur structure, les aspects sociaux comme l'établissement de profils et de politiques de conservation fondés sur la collectivité et la technologie, comme la conception de cadres de fonds de données et d'outils logiciels qui ajoutent de la valeur et une fonctionnalité aux ensembles de données.

---

<sup>5</sup> Groupe de travail sur la stratégie des données de recherche, p. 13

<sup>6</sup> *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. p. 20



### **ODESI et le the Conseil des bibliothèques universitaires de l'Ontario**

<http://odesi.ca>

<odesi> est un outil basé sur le Web qui permet l'exploration, l'extraction et l'analyse des données. <odesi> est un projet subventionné conjointement par le Conseil des bibliothèques universitaires de l'Ontario (CBUO) et OntarioBuys qui offre aux chercheurs universitaires un accès sans précédent à un grand nombre d'ensembles de données. Le projet vise avant tout les ensembles de données de Statistique Canada, les fichiers de données historiques de Gallup Canada et d'autres entreprises de sondage, les fichiers du domaine public comme les sondages sur les élections nationales au Canada et certains fichiers de l'Inter-University Consortium for Political and Social Research (ICPSR).

Les recherches sur les variables (questions de sondage) est possible sur des milliers d'ensembles de données dans un nombre croissant de collections. Les chercheurs qui utilisent <odesi> peuvent effectuer des mises en tableau et des analyses en ligne simples et télécharger la plupart des ensembles de données dans des logiciels statistiques pour de plus amples analyses. Les ensembles de données sont téléchargés par la plateforme Scholars Portal qui est déjà bien connue des étudiants et des chercheurs des universités de l'Ontario.

### **DataCite**

[www.datacite.org/](http://www.datacite.org/)

Un groupe de bibliothèques de recherche et de fournisseurs de renseignements techniques, y compris l'ICIST, qui ont établi un partenariat dans le but d'améliorer l'accès aux données de recherche sur l'Internet. Cette initiative a pour objectif de mettre sur pied un organisme sans but lucratif qui permet aux organisations d'enregistrer des ensembles de données de recherche et leur attribuer des identificateurs permanents pour que les ensembles de recherche puissent être manipulés comme objets scientifiques indépendants que l'on peut citer.

En premier lieu, DataCite doit promouvoir l'utilisation d'identificateurs d'objets numériques (ION) pour les ensembles de données. Un ION est utilisé pour citer les ressources électroniques et s'y relier (le texte aussi bien que les données de recherche de même que d'autres types de contenu). L'ION est différent des autres systèmes de référence utilisés en général sur l'Internet, comme URL, étant donné qu'il est relié en permanence à l'objet lui-même et non seulement à l'endroit où se trouve l'objet. La vision à long terme du partenariat consiste à appuyer les chercheurs en leur fournissant des méthodes qui leur permettent de trouver, identifier et citer les ensembles de données de recherche en toute confiance.

### **Archivage et conservation**

Pour qu'on puisse réutiliser les données de recherche et/ou les réadapter, elles doivent être intégrées dans un environnement institutionnel durable soutenu par un entrepôt de données numériques. La conservation des données de recherche nécessite la gestion active des données pendant tout leur cycle de vie qui comporte des activités comme « l'évaluation, la sélection, le dépôt ou l'intégration des données dans un entrepôt, en assurant leur authenticité, en gérant la collection de données et de métadonnées, en rafraîchissant les supports numériques et en transférant les données vers de nouveaux supports numériques<sup>7</sup> ». L'analyse des lacunes de 2008 a permis de découvrir que les données de recherche au Canada ne sont pas systématiquement conservées. Elle indique que la « majeure partie des données recueillies grâce à la recherche ne sont pas enregistrées dans des entrepôts de données et peu d'entrepôts sinon aucun n'ont une pleine capacité de conservation comme le définit l'état du dépôt numérique fiable<sup>8</sup> ». Bien que les données de certaines disciplines sont recueillies par des organismes nationaux, cela

---

<sup>7</sup> Groupe de travail sur la stratégie des données de recherche. p. 3

<sup>8</sup> Groupe de travail sur la stratégie des données de recherche. p. 7



## Comblent les lacunes de données de recherche : Examen de nouveaux services de bibliothèques

représente seulement une petite proportion des ensembles de données créés au moyen des activités de recherche au Canada.

La communauté des bibliothèques s'intéresse de plus en plus à s'engager envers la conservation des données de recherche. Bon nombre d'activités de conservation des données découlent actuellement des initiatives concernant les dépôts institutionnels (DI). Toutefois, il y a des limites au niveau d'interopérabilité possible à l'aide des logiciels actuels de dépôt institutionnel. Les plateformes de DI actuelles n'ont pas encore la fonctionnalité requise pour étiqueter les données au niveau de l'élément, chose qui est nécessaire pour l'interopérabilité et la réutilisation des ensembles de données. De plus, comme les données de recherche sont très hétérogènes, il est peu probable qu'un seul entrepôt de données puisse recueillir toute la gamme des types de données créés par un seul établissement d'enseignement supérieur. Dans son document intitulé *No Brief Candle*, Rick Luce, directeur de Emory University, recommande qu'au départ, les bibliothèques de recherche pourraient davantage contribuer en aidant à conserver des ensembles de données de petite échelle qui découlent des travaux effectués par des groupes de recherche locaux ou de recherche spécifique à un domaine.

### Exemples :

#### **Cueillette de données à l'aide d'un dépôt institutionnel au MIT**

<http://pledge.mit.edu>

Le projet PLEDGE du MIT vise à mettre au point un mécanisme qui permet d'archiver, de conserver et d'assurer l'accès à des données à l'aide du logiciel d'entreposage DSpace. Plus précisément, on utilise les études du MIT sur Dspace qui se trouvent dans le Réseau Dataverse de l'Institute for Quantitative Social Science (IQSS). Les bibliothèques aident le corps enseignant à préparer les données et fournissent de l'aide avec le dépôt de données. Face à ce qui précède, les chercheurs du MIT ont trois options principales, DSpace, le Centre de données Harvard-MIT (qui fournit le réseau Dataverse IQSS personnalisé) et les archives de données de l'Inter-University Consortium for Political and Social Research (ICPSR). Bien que le personnel bibliothécaire fait la promotion de toutes ces options de dépôt, il s'intéresse aussi à fournir un moyen de rendre la recherche effectuée par le corps enseignant disponible dans tous les trois systèmes et ainsi surmonter le problème des utilisateurs finaux qui doivent chercher à différents endroits pour trouver des données.

#### **Réseau de dépôts de données du projet eCrystals Federation**

[http://wiki.ecrystals.chem.soton.ac.uk/index.php/Main\\_Page](http://wiki.ecrystals.chem.soton.ac.uk/index.php/Main_Page)

Ce projet permettra d'établir une base solide de dépôts de données sur la cristallographie pour un groupe international de 14 sites partenaires, avec des métadonnées recueillies par un certain nombre de services combinatoires. Les partenaires doivent collaborer afin d'harmoniser les profils d'application des métadonnées provenant des dépôts qui fonctionnent sur différentes plateformes (EPrints, DSpace, Fedora et ReciprocalNet); examiner les problèmes de combinaison qui surviennent de la cueillette des données provenant des dépôts dans les environnements d'information en réseau des autres pays (UE, É.-U. et Australie), permettre à la Fédération des archives institutionnelles d'interopérer avec les archives-sujets internationales (IUCr et CCDC) et d'autres tiers cueilleurs, s'occuper des secteurs de conservation, de préservation et de provenance des données en devenant membre d'une fédération opérationnelle et, par la prestation d'un banc d'essai pour les outils DRAMBORA, assurer une approche de promotion et, de formation coordonnée en vue de la mise en œuvre des dépôts.

#### **Projet DISC-UK DataShare**

<http://www.disc-uk.org/datashare.html>

Ce projet, dirigé par EDINA et l'Edinburgh University Data Library, qui compte des partenaires à la University of Southampton et Oxford University, a fait progresser la prestation actuelle des services de dépôt pour recevoir des ensembles de données au Royaume-Uni. En collaborant avec les intervenants de

quatre universités et à l'interne avec les collègues déjà engagés à la gestion de l'accès ouvert des dépôts pour les manuscrits d'articles scientifiques (*e-prints*), ce partenariat a mis en place et mis à l'essai de nouveaux modèles de partage et d'archivage des données dans les établissements de recherche du Royaume-Uni. En appuyant, dans les quatre établissements partenaires, les chercheurs qui souhaitent partager des ensembles de données sur lesquels sont fondés les résultats de la recherche, ce réseau de dépôts de données institutionnel a conçu un modèle spécialisé visant le dépôt des « ensembles de données orphelins » qui, jusqu'à maintenant, ne faisait l'objet ni d'archives/centres/réseaux centralisés de données de sujets-domaines ni de dépôts institutionnels basés sur des *e-prints*.

### **Environnements de recherche virtuelle**

La recherche virtuelle se décrit comme une vraie méthode de recherche, alimentée par des capacités en réseau et de vastes quantités de données. La recherche virtuelle modifie en profondeur la façon dont les chercheurs exécutent leur travail, les outils qu'ils utilisent, les types de problèmes qu'ils doivent régler et la nature des documents et des publications qui résultent de leur recherche. Les environnements de recherche virtuelle (ERV) font désormais partie de la cyberinfrastructure. Ils mettent en place un cadre de ressources qui appuie les activités de recherche sous-jacentes et se définissent comme un ensemble d'outils, de systèmes et de procédés en ligne qui fonctionnent les uns avec les autres afin de faciliter ou améliorer la recherche à l'intérieur et à l'extérieur des limites institutionnelles<sup>9</sup> ».

La participation des bibliothèques à la recherche virtuelle et aux ERV est plutôt récente. Comme le mentionnait Cliff Lynch, les ERV « peuvent sembler ne pas s'adapter à la tradition plutôt passive des bibliothèques<sup>10</sup> ». De plus, les ERV sont souvent utilisés entre institutions et appuient la recherche collective parmi les chercheurs dispersés sur le plan géographique, ce qui soulève des difficultés pour la participation des bibliothèques. Si des bibliothèques peuvent surmonter ces obstacles, elles seraient bien placées pour contribuer au perfectionnement des ERV sur les campus.

### **Exemples :**

#### **myExperiment.org**

<http://myexperiment.org>

L'environnement de recherche virtuelle myExperiment appuie le partage des objets de recherche utilisés par les scientifiques, comme les flux de travaux scientifiques. Pour les chercheurs, il s'agit là d'une infrastructure sociale qui encourage le partage, et d'une plateforme qui permet d'effectuer la recherche au moyen d'interfaces utilisateurs bien connues. Pour les développeurs, c'est un environnement ouvert, extensible et participatif. Le logiciel qui appuie myexperiment.org peut être téléchargé pour que les utilisateurs puissent exécuter leur propre occurrence de myExperiment.

#### **Dépôt de mise à disposition des données DataStaR de l'Université Cornell**

<http://datastar.mannlib.cornell.edu/>

DataStaR est conçu comme dépôt de mise à disposition des données. Il fonctionne comme dépôt temporaire pour les ensembles de données de recherche (à toute étape d'achèvement). Les chercheurs qui présentent leurs données ont accès à un espace de travail qui leur permet d'assurer un partage contrôlé avec certains collègues (ou avec le public si le chercheur le souhaite) et qui assure également un stockage et une sauvegarde à distance pour les données. Une fois qu'elles respectent les exigences de dépôt dans

---

<sup>9</sup> Définition tirée du groupe consultatif e-Infrastructure OSI du Royaume-Uni : [www.nesc.ac.uk/documents/OSI/vrc.pdf](http://www.nesc.ac.uk/documents/OSI/vrc.pdf)

<sup>10</sup> Lynch, C. 2008. p. 1

une archive ou un dépôt de données disciplinaires externes, les données peuvent passer du dépôt de mise à disposition [DataStaR] à la publication. Les chercheurs qui stockent leurs données dans le dépôt de mise à disposition ont également l'option de présenter des données « prêtes pour la publication » dans le DI de l'Université Cornell, eCommons.

**Islandora : Environnement de recherche virtuelle à l'Université de l'Île-du-Prince-Édouard**

<http://islandora.ca/>

Islandora est un projet en exploitation libre en cours d'exécution à la bibliothèque Robertson. Islandora combine les logiciels Drupal et Fedora pour mettre en place un système solide de gestion des actifs numériques que l'on peut utiliser lorsqu'on a besoin de collaboration et de gérance des données numériques, à court et à long terme. Islandora assure un service essentiel grâce à son environnement de recherche virtuelle (ERV), un site Web spécial utilisé par les chercheurs de l'UPEI et ailleurs pour gérer les données de recherche. Un nombre croissant d'établissements partout dans le monde utilisent Islandora pour leurs propres systèmes de dépôt. L'ERV est un ensemble de sites personnalisés Islandora utilisés par les chercheurs de l'UPEI et d'ailleurs pour gérer les données de recherche.

## **Prochaines étapes**

Toute solution visant la gérance des données au Canada doit traiter des difficultés d'un ensemble du cycle de vie des données de recherche, tout en tenant compte de la grande diversité des ensembles de données de toutes les disciplines. Évidemment, aucun établissement ne peut régler ce type de problèmes seul. Une approche globale nécessite que tous les intervenants, c'est-à-dire les chercheurs, les organismes de financement, les ministères gouvernementaux, les universités et les bibliothèques, collaborent afin de définir les rôles et responsabilités et poursuivent des objectifs d'gérance de plus haut niveau qui sont importants pour l'ensemble de la collectivité<sup>11</sup>. Cela dit, les bibliothèques peuvent commencer à poser les jalons d'une approche nationale mieux coordonnée et plus complète.

Dans un récent article au sujet des programmes de conservation des données aux États-Unis, Tyler Walters, du Georgia Institute of Technology décrit la situation aux États-Unis. « La conservation des données de recherche scientifique dans les universités américaines est un récit de personnes entreprenantes et de progrès cumulatifs. Un petit nombre de bibliothèques et de centres de données qui voient la possibilité de devenir des centres de gestion de l'information numérique, ont pris des mesures entrepreneuriales pour aller au-delà de leurs actifs numériques traditionnels et passer à la gestion des données de recherche scientifique et savante<sup>12</sup>. De même au Canada, les bibliothèques qui s'intéressent actuellement à mettre en œuvre des services de données de recherche devront sans doute le faire sans l'appui de politiques visant l'ensemble des établissements ni de programmes de financement gouvernemental ciblés. Toutefois, compte tenu de la tendance internationale croissante vers la gérance des données, ces programmes pourraient être le ferment qui donne naissance à des initiatives officielles à l'avenir.

En fin de compte, les lacunes qui existent en matière de gérance des données de recherche au Canada présentent une occasion pour les bibliothèques de prendre plus de place dans l'infrastructure de recherche virtuelle en pleine évolution. Le présent examen a trouvé de nombreuses possibilités pour les bibliothèques de participer davantage dans le secteur des domaines de recherche. Certaines recommandations pour les bibliothèques qui envisagent de procéder ainsi figurent ci-dessous.

## **Évaluation des besoins**

Pour mettre en place des services efficaces, les bibliothèques doivent avoir une bonne idée de ce que veulent leurs chercheurs et de la nature des actifs de données qui se trouvent dans l'établissement. Ce type de renseignements peut être recueilli au moyen d'une analyse des besoins. En 2007, le Joint Information Systems Committee du Royaume-Uni a financé la mise au point d'un cadre de vérification des données (CVD) afin d'aider à évaluer les besoins en matière de données. Ce cadre est, pour les organismes, un moyen de définir, de trouver, de décrire et d'évaluer les façons dont ils gèrent leurs données de recherche. Il regroupe un ensemble de méthodes et un outil en ligne qui permettent aux vérificateurs des données de recueillir les renseignements pertinents. C'est un outil pratique qui peut être utilisé ou adapté par les bibliothèques canadiennes dans le but d'évaluer les besoins des chercheurs et d'élaborer des services de données qui tiennent compte de ces besoins.

## **Perfectionnement des compétences**

Bien que certaines activités comme la conscientisation et la promotion peuvent ne pas nécessiter du personnel de bibliothèque qu'il acquière d'importantes nouvelles connaissances, il est évident que les bibliothèques devront investir dans le perfectionnement des compétences du personnel qui s'occupe de services de données de recherche plus complexes. Les compétences fondamentales des bibliothécaires dans le domaine de la gestion des données comprennent : convertir les données dans de nouveaux

---

<sup>11</sup> Groupe de travail sur la stratégie des données de recherche. p. 7

<sup>12</sup> Walters, T. p. 1

formats, concevoir des ontologies, des structures hiérarchiques et des dictionnaires analogiques interactifs, produire des métadonnées; organiser et préserver les données et mettre au point des outils d'accès et de découverte. Un certain niveau de compétence dans certains domaines est également souhaitable puisque, idéalement, les spécialistes des données comprendront les besoins des disciplines particulières.

Par le passé, les compétences en gestion des données ont souvent été acquises en milieu de travail, au fil du temps, grâce aux scientifiques locaux qui produisaient des données, ou au moyen d'une collaboration avec le personnel enseignant sur les projets de recherche. Toutefois, la formation en milieu de travail peut ne pas suffire pour les bibliothèques qui visent à mettre en œuvre de vastes services de données de recherche. Il existe un certain nombre de programmes d'études de deuxième cycle accrédités et de programmes de formation et d'enseignement professionnel qui sont disponibles pour les bibliothécaires à l'extérieur du Canada. Par exemple, le Data Curation Education Program (DCEP) à la Graduate School of Library and Information Science de l'Université de l'Illinois d'Urbana-Champaign, prépare les étudiants à travailler dans le domaine de la conservation des données. On y offre également des stages d'été pour donner une formation aux bibliothécaires qui travaillent. Il existe un certain nombre d'autres occasions de formation aux États-Unis et au Royaume-Uni. On s'attend également à ce qu'un plus grand nombre de ce type de programmes soient disponibles compte tenu de la demande croissante. Au Canada, l'Association des bibliothèques de recherche du Canada (ABRC), le Groupe de travail sur la stratégie des données de recherche du Canada et la bibliothèque de l'Université Carleton commanditent un certain nombre d'ateliers sur les données de recherche pour les bibliothécaires partout au Canada.

### **Nouveaux modèles organisationnels**

Selon un rapport commandé en 2008 par JISC au Royaume-Uni, l'appui à la recherche axée sur les données nécessitera un repositionnement stratégique des bibliothèques<sup>13</sup>. Les bibliothèques de recherche ont par le passé été structurées et dotées en fonction des disciplines. Par contraste, la recherche virtuelle favorise les approches multidisciplinaires. Selon Rick Luce, la recherche virtuelle nécessite souvent des équipes virtuelles qui se forment dynamiquement au cours des étapes de planification initiale d'un projet de recherche, travaillent à un projet puis se transforment en autre chose lorsqu'une présence moins intensive est nécessaire. Cela donne lieu à des structures de dotation souples et un modèle structural plus dynamique que la méthode actuelle d'attribuer des liens entre départements ou entre sujets<sup>14</sup>. Pour Luce, l'avenir comporte des bibliothèques qui peuvent mettre sur pied des équipes de gestion de l'information polyvalentes à la volée et intégrer les bibliothécaires aux équipes ou départements de recherche.

Les changements de structures de bibliothèques de recherche devraient être entraînés par les besoins des collectivités de recherche avec lesquelles elles collaborent. Cela nécessitera un changement d'orientation de la fourniture de produits vers le soutien des procédés. Étant donné que le cycle de vie des données commence tôt dans le processus scientifique, les nouveaux services de bibliothèque doivent également être élaborés dès les premières étapes d'un projet de recherche, et il faudra également mettre en place des services en aval pour la préservation et la réutilisation des données<sup>15</sup>.

### **Modèles de coûts et exécution d'une analyse de rentabilisation**

Les coûts de la plupart des services de données de recherche sont peu importants. Selon Rick Luce, « les mécanismes largement décentralisés et peu uniformes de création des données de tous les types et formats imaginables rendent l'évaluation des coûts de conservation et de préservation des données plutôt compliquée »<sup>16</sup>. En 2008, le JISC a financé une étude visant à évaluer les coûts de préservation des

---

<sup>13</sup> Swan, A. p. 1

<sup>14</sup> Luce, R. p. 1

<sup>15</sup> Gold, A. p. 1

<sup>16</sup> Luce, R. p. 1

données de recherche. *Keeping Research Data Safe : A Cost Model and Guidance for UK Universities* propose un fondement méthodologique qui permet d'évaluer les coûts de la cueillette de données de recherche. Cette étude illustre que la majeure partie des coûts sont ceux qui se rapportent au personnel (environ 85 %) et à l'équipement (environ 13 %). On recommande que les bibliothèques qui s'occupent actuellement de concevoir un modèle de coût pour les services de données consultent ce rapport.

L'exécution d'une analyse de rentabilisation pour les services de données peut être assez difficile. On ne peut pas toujours relier directement les avantages à la production de recherche améliorée à l'intérieur d'une unité particulière de recherche, mais les résultats sont surtout utilisés par la collectivité de recherche en général. Toutefois, l'un des avantages indirects est le rehaussement de l'image d'un établissement partout dans le monde. D'autres avantages, qui font l'objet de discussions plus détaillées dans le document d'accompagnement intitulé *Les données de recherche : un potentiel insoupçonné*, comprennent la maximisation de l'investissement dans les collections, la réduction du dédoublement des coûts et des collections de données, l'accroissement de la transparence des enregistrements scientifiques et la possibilité de nouvelles découvertes scientifiques.

## References

Gold, A. (2007). 'Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries'. *D-Lib Magazine* 13, (9/10). 2007.

[Available at: [www.dlib.org/dlib/september07/gold/09gold-pt2.html](http://www.dlib.org/dlib/september07/gold/09gold-pt2.html)]

*Harnessing the Power of Digital Data for Science and Society*. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Sciences and Technology Council, January 2009.

Joint Task Force on Library Support for E-Science. *Agenda for Developing E-Science in Research Libraries*. Association of Research Libraries. Association of Research Libraries, November 2007. [Available at: [http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf)]

*Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. National Science Foundation, September 2005. [Available at: <http://www.nsf.gov/pubs/2005/nsb0540/>]

Luce, Richard E. 'A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries'. In *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. CLIR Report 142, August 2008. [Available at: <http://www.clir.org/pubs/reports/pub142/luce.html>]

Lynch, C. The Institutional Challenges of Cyberinfrastructure and E-Research. *EDUCAUSE Review*, 43(6). 2008. [Available at: [www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/TheInstitutionalChallengesofCy/163264](http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/TheInstitutionalChallengesofCy/163264)]

Lynch, Clifford. "Where Do We Go From Here? The Next Decade for Digital Libraries". *D-Lib Magazine* 11(7/8), July/August 2005. [Available at: [www.dlib.org/dlib/july05/lynch/07lynch.html](http://www.dlib.org/dlib/july05/lynch/07lynch.html)]

Lyon, L. *Dealing with data: Roles, rights, responsibilities and relationships: Consultancy report*. UKOLN, University of Bath, UK. June 17, 2007. [Available at: [www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing\\_with\\_data\\_report-final.pdf](http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf)]

Martinez-Urbe, L. *Findings of the scoping study interviews and the Research Data Management Workshop: Scoping digital repository services for research data management*. University of Oxford, UK. May 27, 2008. [Available at: [www.ict.ox.ac.uk/odit/projects/digitalrepository/](http://www.ict.ox.ac.uk/odit/projects/digitalrepository/)]

Neil Beagrie, Julia Chruszcz, and Brian Lavoie. Keeping Research Data Safe: A Cost Model and Guidance for UK Universities. Joint Information Systems Committee, 2008. [Available at: <http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx>]

Groupe de travail sur la stratégie des données de recherche. *Gérance des données de recherche au Canada : une analyse des écarts*. Institut canadien de l'information scientifique et technique [disponible à l'adresse : <http://data-donnees.gc.ca/docs/AnalyseÉcarts.pdf>]



## Comblent les lacunes de données de recherche : Examen de nouveaux services de bibliothèques

Shearer, Kathleen. *Les données de recherche : un potentiel insoupçonné*. Association des bibliothèques de recherche du Canada, 2009 [disponible à l'adresse : [www.carl-abrc.ca/about/working\\_groups/pdf/data\\_mgt\\_toolkit.pdf](http://www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf)]

Swan, Alma and Sheridan Brown. *Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs*. Joint Information Systems Committee, July 2008. [Available at: <http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>]

*To Stand the Test of Time: Long-term Stewardship of Data Sets in Science and Engineering*. A Report of the ARL/NSF Workshop on Long-Term Stewardship of Digital Data Collections, September 26-27, 2006. Association of Research Libraries [Available at: [www.arl.org/pp/access/nsfworkshop.shtml](http://www.arl.org/pp/access/nsfworkshop.shtml)]

Wilbanks, John. "I Have Seen the Paradigm Shift, and It Is Us". *The Fourth Paradigm: Data Intensive Scientific Discovery*. Microsoft Research, 2009. pg. 209-214. [Available at: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>]

Walters, Tyler O. "Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example". *International Journal of Digital Curation*, 4 (3), 2009. [Available at: <http://www.ijdc.net/index.php/ijdc/article/view/136>]

Wusteman, Judith. Virtual Research Environments: What Is the Librarian's Role? *Journal of Librarianship and Information Science*, 40 (2) June, 2008.