

# Tirer profit des données massives

Réponse de l'ABRC au document de consultation Tirer profit des données massives : les organismes subventionnaires tiennent une consultation sur la recherche numérique au Canada

*13 décembre 2013*



**Canadian Association  
of Research Libraries**

**Association des bibliothèques  
de recherche du Canada**

## **Qui nous sommes**

L'Association des bibliothèques de recherche du Canada (ABRC) représente les vingt-neuf plus grandes bibliothèques universitaires du pays. L'amélioration de la recherche et de l'enseignement supérieur sont au cœur de notre mission. L'ABRC favorise l'efficacité et la pérennité de la communication savante, ainsi que les politiques publiques qui permettent l'accès le plus large possible à l'information savante.

## **Vue d'ensemble**

L'Association des bibliothèques de recherche du Canada (ABRC) remercie le Conseil de recherches en sciences humaines (CRSH), les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie (CRSNG) et la Fondation canadienne pour l'innovation (FCI), agissant en collaboration avec Génome Canada – le « groupe 3+ » – d'avoir consulté la grande collectivité des intervenants sur une éventuelle politique concernant la gérance des données de recherche par leur document intitulé *Tirer profit des données massives : les organismes subventionnaires tiennent une consultation sur la recherche numérique au Canada*.

Les membres de l'ABRC sont solidement engagés dans les questions de gérance des données de recherche au sein de notre association, de nos établissements et de nos consortiums régionaux de bibliothèques et par l'exercice de rôles très actifs dans Données de recherche Canada (DRC) et le Conseil de direction pour l'infrastructure numérique (CDIN). L'ABRC a eu l'occasion de contribuer aux réponses de DRC et du CDIN au document de consultation, et nous savons que d'autres organismes connexes présentent des instances également. Dans cette présentation de l'ABRC, nous traitons des enjeux dans l'optique particulière des bibliothèques de recherche dans le contexte national.

Les membres de l'ABRC abordent ces questions dans la perspective des chercheurs avec lesquels nous travaillons au quotidien. Ces chercheurs veulent acquérir et utiliser des données par de nouveaux moyens computationnels, ils veulent planifier la gestion efficace des données générées dans les nouveaux projets de recherche, et ils veulent présenter et préserver leurs données pour plus tard. Dans cette perspective de chercheur, l'ABRC est heureuse de formuler les recommandations suivantes, qui sont expliquées plus en détail ci-après, en réponse au document de consultation :

1. Travailler avec les collectivités d'intervenants clés pour faire de la sensibilisation à l'importance de la gestion des données de recherche
2. Proposer des lignes directrices pour des éléments particuliers dans les plans de gérance des données
3. Exploiter les points forts des bibliothèques de recherche en matière de gérance pour élaborer des services de gestion des données de recherche
4. Établir des processus de certification de la préservation pour opérationnaliser des dépôts de données appropriés
5. Faciliter la coordination des initiatives existantes et y prendre appui

6. Mettre au point des mécanismes de financement appropriés pour appuyer la gestion des données de recherche tout au long du cycle de vie des données
7. Investir dans la formation et le perfectionnement des compétences dans l'ensemble du cycle de vie de la gestion des données de recherche

L'ABRC félicite le groupe 3+ de son intention d'attirer toute l'attention sur la gérance des données de recherche au Canada. Comme vous le savez, les données de recherche représentent un précieux avoir financé sur les fonds publics qui, bien géré et partagé équitablement, peut être exploité plus à fond et mener à de nouvelles découvertes et à une plus grande innovation. En outre, les résultats de la recherche ne peuvent être validés ni reproduits à moins que les données de recherche qui les sous-tendent ne soient accessibles et compréhensibles pour d'autres. Les lacunes de la capacité du Canada de gérer les données de recherche sont déjà bien documentées. Cela dit, le Canada a aussi certains points forts qui peuvent être mis dans l'équation de cette initiative, comme diverses activités de gestion des données de recherche par discipline, ainsi que les activités de collaboration des bibliothèques de recherche à la gestion d'une gamme de produits de recherche essentiels. D'autres actions du groupe 3+ aideront à mettre le Canada au diapason des autres pays développés qui ont déjà commencé à se donner l'infrastructure et les politiques nécessaires pour appuyer le partage à grande échelle et la conservation des données de recherche.

## ***Recommandations***

### **Mise en place d'une culture de gouvernance**

#### **Recommandation 1 : Travailler avec les collectivités d'intervenants clés pour faire de la sensibilisation à l'importance de la gestion des données de recherche**

La gestion des données de recherche (GDR) en contexte universitaire met en cause toute une gamme d'intervenants, comme les administrateurs, les chercheurs, les bibliothèques et les services de TI, qui ont tous un important rôle à jouer. Les vice-présidents à la recherche, assistés de leurs bureaux d'appui à la recherche, assureront le leadership sur le campus avec les bibliothécaires universitaires et les dirigeants principaux de l'information et il est important que le groupe 3+ travaille avec ces groupes. En outre, il sera crucial de faire connaître les difficultés que vit la collectivité de la recherche en s'assurant la participation des sociétés scientifiques et savantes et des associations de professeurs. Le plan du groupe 3+ pour le développement de la capacité devrait comprendre un programme d'appui à la promotion et à l'éducation sur les campus pour aider les établissements à faire de la sensibilisation et à développer les compétences. Ainsi, ce type de programme pourrait comprendre des équipes d'experts qui se rendraient disponibles pour conseiller les établissements sur divers éléments de la gouvernance de la GDR et du soutien de la GDR.

#### **Recommandation 2 : Proposer des lignes directrices pour des éléments particuliers des plans de gérance des données**

Exiger des plans de gérance des données dans le cadre des demandes de financement est un bon premier pas qui incitera les chercheurs à songer à la façon dont seront gérées les données qu'ils produisent. Compte tenu des niveaux variables de sensibilisation à la valeur de la GDR à long terme parmi les chercheurs et de l'état fragmenté de l'appui à l'échelle du pays, les exigences initiales pour les plans de gérance des données pourraient être relativement vastes et flexibles.

Par contre, il y aurait lieu de resserrer les exigences au fil du temps, au fur et à mesure que la collectivité comprendra et acceptera les besoins et que le Canada aura développé plus à fond ses capacités et son infrastructure. Les expériences des bailleurs de fonds ailleurs ont démontré que toutes les exigences de politique pour les plans de GDR devraient préciser les éléments particuliers à couvrir dans le plan. La collectivité des bibliothèques de recherche a déjà commencé à travailler avec les chercheurs à des plans de GDR et l'ABRC, avec Données de recherche Canada, serait heureuse de formuler des recommandations précises sur les éléments désirables d'un plan de GDR.

### **Recommandation 3 : Exploiter les points forts des bibliothèques de recherche en matière de gérance pour élaborer des services de gestion des données de recherche**

Les bibliothèques de recherche ont une longue tradition de gérance. Nos valeurs de base comprennent la préservation, l'accès à l'information, et la protection des renseignements personnels. Nos bibliothécaires sont compétents non seulement dans les pratiques de saine gestion de l'information, mais encore en enseignement et en recherche. Certaines bibliothèques de recherche, et en particulier celles des grandes universités, ont déjà des bibliothécaires avec expertise en gestion des données qui peuvent offrir de l'aide aux chercheurs pour la planification de la gestion des données. L'ABRC elle-même a donné une certaine formation en GDR au personnel des bibliothèques, en offrant un institut de quatre jours en 2013, et en offrira probablement un deuxième, plus spécifique ou avancé en 2014. Cependant, ces programmes pourraient être élargis s'il y avait plus de financement disponible à cette fin. Par ailleurs, il subsiste d'importants écarts dans les services et le groupe 3+ devrait travailler avec les bibliothèques de recherche pour développer la capacité dans ce domaine. (Voir aussi la recommandation 7.)

### **Recommandation 4 : Établir des processus de certification de la préservation pour opérationnaliser les dépôts de données appropriés**

Un des éléments du plan de gérance des données que propose le groupe 3+ comprend des « dépôts de données appropriés ». Il sera important d'établir une compréhension commune des dépôts de données et de faire comprendre qu'il ne s'agit pas simplement d'espace d'entreposage pour l'analyse computationnelle ou d'autre entreposage à court terme. En particulier, la préservation numérique et la conservation des données doivent faire partie intégrante de la gouvernance des données. Les bibliothèques de recherche ont une expertise et une expérience considérables à offrir à cet égard (voir la recommandation 5). Le groupe 3+ et l'ABRC, avec Données de recherche Canada, doivent se concerter pour formuler des critères appropriés et un processus de certification pour les dépôts. L'Alliance des données de recherche (ADR) a lancé un groupe chargé de définir des critères globaux de certification des dépôts de données qui regroupera diverses initiatives régionales de certification. Ces travaux pourraient constituer un cadre approprié pour la détermination de ces paramètres ici au Canada.

### **Coordination de la mobilisation des intervenants**

#### **Recommandation 5 : Faciliter la coordination des initiatives existantes et y prendre appui**

Les grandes composantes de l'infrastructure de GDR existent déjà dans une certaine mesure, mais il faudra de la coordination et des ressources pour « souder ensemble » les autres composantes et corriger les lacunes qui subsistent. Tout en convenant que la coordination sera

toujours une bonne chose, nous ferions valoir que, à ce stade-ci, l'enjeu n'est pas le doublement ou le chevauchement des intérêts parmi les fournisseurs de services d'infrastructure de recherche numérique (IRN); les lacunes bien connues d'infrastructure (au moins au niveau de la GDR) sont un bien plus grand problème.

En particulier, le groupe 3+ devrait consulter à la fois le Conseil de direction pour l'infrastructure numérique (CDIN) et Données de recherche Canada (DRC), qui sont très représentatifs de nombreux intervenants à l'échelle du pays. DRC a une articulation et une force particulières dans l'aspect GDR de l'IRN. Quant à l'entreposage des données à long terme, Calcul Canada pourrait pouvoir consacrer plusieurs pétaoctets d'espace-disque, qu'il pourrait être possible de structurer pour permettre un régime de préservation actif.

La préservation suppose non seulement l'aménagement de l'espace d'entreposage approprié, mais également la gestion à long terme des données. La collectivité des bibliothèques de recherche a travaillé aux premières lignes de la préservation numérique au Canada. Le Scholars Portal du Conseil des bibliothèques universitaires de l'Ontario (CBUO), qui loge à l'University of Toronto, a élaboré le premier « référentiel numérique de confiance » certifié au Canada. Cette certification a été délivrée en vertu de la norme ISO 16363 en février 2012 après un audit du dépôt des revues du Scholars Portal. Le dépôt, qui est en place depuis une décennie, renferme plus de 37 millions d'articles de revues. Le CBUO est en voie de créer l'Ontario Digital Library Research Cloud, c'est-à-dire un nuage de recherche de 1,2 Po avec nœuds distribués à l'échelle de la province qui s'intégrera avec les applications de dépôt. Les membres du Council of Prairie and Pacific University Libraries ont de l'expérience du programme LOCKSS, basé aux Stanford University Libraries, qui offre un environnement de préservation numérique pour le contenu numérique. Un réseau LOCKSS pour l'information gouvernementale canadienne a été créé par un groupe de onze bibliothèques universitaires à l'automne 2012. De même, plusieurs bibliothèques universitaires (Alberta, British Columbia et Simon Fraser) dans l'Ouest canadien collaborent à la création d'un environnement de préservation numérique en exploitation libre, le système à base de normes Archivematica.

Nous devrions signaler également l'existence de plusieurs autres initiatives basées dans les bibliothèques qui présentent de l'intérêt pour la GDR, et notamment DataCite, dont le nœud canadien est logé à l'Institut canadien de l'information scientifique et technique du Conseil national de recherches du Canada (CNRC-ICIST), qui offre des ION (identificateurs d'objet numérique) pour les ensembles de données archivées, et le Réseau canadien d'information polaire (RCIP), qui a mis au point avec succès un régime d'entreposage distribué de données et de préservation active des données. En outre, plusieurs bibliothèques universitaires individuelles archivent des ensembles de données dans des archives numériques et offrent des services qui rejoignent d'autres aspects importants de la GDR : éducation, formation, conseils et aide aux chercheurs.

Pour coordonner les diverses initiatives basées dans les bibliothèques et d'élargir le soutien pour la GDR dans les établissements et à l'échelle des régions, l'ABRC parraine la création d'un « réseau » de ces initiatives, dont la forme et les buts précis sont actuellement en discussion. Le rôle des consortiums régionaux de bibliothèques universitaires et du Réseau canadien de documentation pour la recherche (RCDR) seront une considération importante dans cette approche puisque leurs membres représentent collectivement la totalité des bibliothèques universitaires publiques à l'échelle du Canada. Le RCDR, qui a su créer une plateforme

nationale, pourrait être un véhicule approprié pour la coordination administrative et la distribution des fonds aux nœuds régionaux d'un réseau.

Alors que ces initiatives basées dans les bibliothèques contribueront à répondre à certains besoins de GDR pour l'entreposage de préservation, la formation et les services à l'échelle du Canada, on ne prévoit pas qu'elles sauront répondre à plus qu'une faible fraction de l'ensemble des besoins; les bibliothèques universitaires, directement ou par l'ABRC, seront heureuses de collaborer et de se faire partenaires avec d'autres organismes capables de contribuer à l'effort national de GDR.

### **Développement de la capacité et paramètres du financement futur**

#### **Recommandation 6 : Mettre au point des mécanismes de financement appropriés pour appuyer la gestion des données de recherche tout au long du cycle de vie des données**

Pour arriver à une « exploitation soutenue » de « centres de classe mondiale spécialisés en gestion des données », il faudra réévaluer les paramètres du financement d'une IRN d'envergure nationale. Le pouvoir de convocation du groupe 3+ sera une importante incitation aux divers intervenants à travailler ensemble, y compris à rallier les provinces à une approche pancanadienne de l'infrastructure numérique. Pour ce qui est du financement des plateformes nationales de recherche, la FCI a réussi, par le passé, à financer une infrastructure canadienne avec les contributions des provinces. Un bon exemple est le financement du Projet canadien de licences de site nationales (PCLSN), qui est devenu le Réseau canadien de documentation pour la recherche (RCDR). Il faudrait envisager une approche semblable du financement pour la gestion des données de recherche.

#### **Recommandation 7 : Investir dans la formation et le perfectionnement des compétences dans l'ensemble du cycle de vie de la gestion des données de recherche**

Un grand défi de capacité est l'acquisition de compétences et de connaissances dans les diverses collectivités d'intervenants. La GDR est complexe et demande du personnel qualifié de tous les secteurs du cycle de vie de la recherche, soit des chercheurs, des spécialistes des données et des bibliothécaires. Le groupe 3+ pourrait vouloir envisager d'investir dans des programmes de formation pour la GDR dans les collectivités d'intervenants. Dans ce contexte, le Digital Curation Centre (DCC) au Royaume-Uni pourrait constituer un modèle intéressant pour le Canada. Le DCC est un centre d'expertise en conservation d'information numérique, qui s'attache à développer la capacité, les capacités et les compétences pour la GDR dans toute la collectivité de recherche en enseignement supérieur au Royaume-Uni. Ce modèle servirait à appuyer les genres de programmes de formation que l'ABRC a lancés et ceux proposés dans certains établissements, et comblerait les lacunes observées par le Sous-comité de l'éducation et de la formation de Données de recherche Canada.

### ***Conclusion***

En conclusion, l'ABRC serait heureuse de participer à des discussions sur ces diverses propositions, qui pourraient trouver grâce aux yeux des intervenants et être développées plus en détail. En général, nous sommes très encouragés par les propositions mises de l'avant dans le document de consultation car elles traduisent la détermination du groupe 3+ à faire des progrès

dans la coordination du développement et le financement de l'IRN et le développement et la viabilisation d'une approche canadienne de la gestion des données de recherche. Encore une fois, nous remercions le groupe 3+ d'avoir consulté le milieu et nous nous réjouissons à la perspective d'aller de l'avant, comme pays, vers une GDR plus poussée et plus efficace.

**Association des bibliothèques de recherche du Canada**

Tél. 613.482.9344 203-309 Cooper  
[info@carl-abrc.ca](mailto:info@carl-abrc.ca) Ottawa (Ont.)  
[www.carl-abrc.ca](http://www.carl-abrc.ca) K2P 0G5

