



Conseil de recherches en  
sciences humaines du Canada

Social Sciences and Humanities  
Research Council of Canada

Canada

Archives nationales  
du Canada

National Archives  
of Canada



## **Consultation nationale sur les archives de résultats de recherche**

Phase un : Rapport d'évaluation  
des besoins

Mai 2001

# Table des matières

<b>Résumé</b>	<b>1</b>
<b>1.0 Introduction</b>	
1.1 Contexte	5
1.2 Processus de consultation	7
1.3 Réponses aux questionnaires	9
1.4 Présentations des intervenants	11
<b>2.0 Lacunes dans l'infrastructure de recherche existante</b>	
2.1 Mandats des Archives nationales du Canada et de la Bibliothèque nationale du Canada	12
2.2 Politique du Conseil de recherches en sciences humaines du Canada en matière d'archivage des données	13
2.3 Services universitaires de traitement des données	13
2.4 Authentification des résultats de recherche	14
2.5 Présence canadienne sur la scène internationale	14
<b>3.0 Besoins d'archivage du milieu de la recherche</b>	
3.1 Données perdues	15
3.2 Occasions perdues	15
3.3 Connaissances et attitudes des chercheurs	16
3.4 Sécurité des résultats de recherche	16
<b>4.0 Avantages d'un service national d'archivage des données</b>	
4.1 Personnes qui en tirent profit	17
4.2 Création de ressources pour l'analyse multidisciplinaire	18
4.3 Contribution à l'enseignement et à la formation en recherche	18
4.4 Mise en valeur des ressources en archivage des données	19
4.5 Préservation et accessibilité des ressources textuelles électroniques	19
4.6 Données ne pouvant être reproduites	19
4.7 Modèles de conception pour les secteurs public et privé	20
4.8 Accessibilité des chercheurs à des ressources coûteuses	20
<b>5.0 Conclusion</b>	<b>21</b>
<b>6.0 Recommandation</b>	<b>21</b>
<b>Annexes</b>	
1) Enquête auprès des chercheurs subventionnés par le CRSH	22
2) Liste des membres des groupes de travail et d'étude	38
3) Liste des présentations des intervenants	40
4) Liste des rapports sur l'archivage des résultats de recherche et autres documents connexes	42

## Résumé

Le présent document fait état de la première phase d'évaluation des besoins en matière de consultation des archives nationales de données. Il est destiné au Conseil d'administration du CRSH et à l'archiviste national du Canada.

Le rapport prouve qu'il y a une lacune importante dans l'infrastructure de recherche canadienne : l'absence d'un service national d'archivage des résultats de recherche. C'est ce qu'on a pu recueillir du milieu universitaire de la recherche en sciences humaines et sociales, c'est-à-dire des archivistes de données et des bibliothécaires d'universités et d'autres établissements privés ainsi que d'une grande variété d'intervenants préoccupés par la conservation et la gestion des documents de recherche numériques.

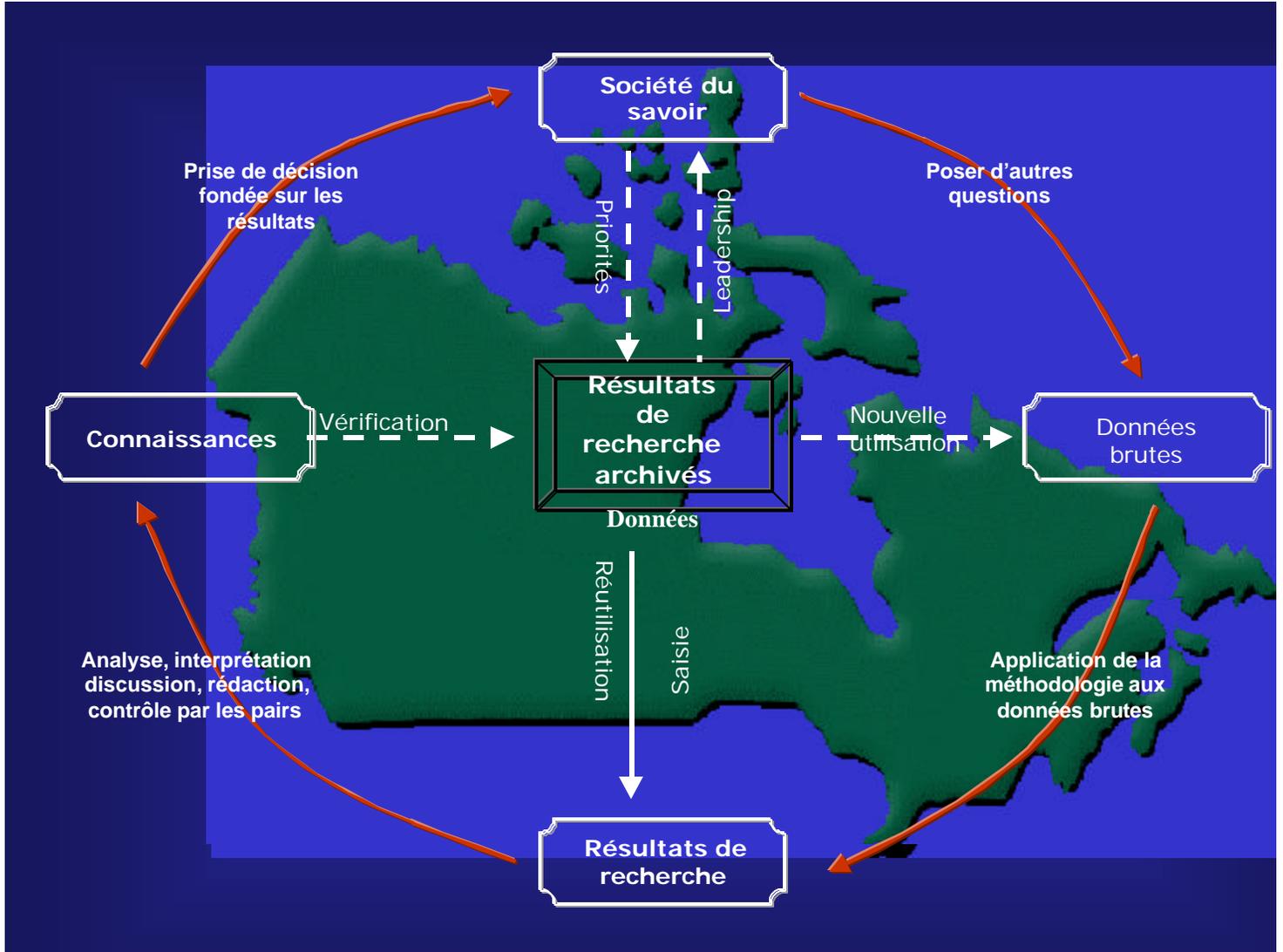
Le CRSH et les Archives nationales (AN) du Canada ont demandé à un groupe de spécialistes d'examiner si les structures et mandats des établissements existants répondent aux besoins d'archivage des résultats de recherche, en quoi consistent ces besoins et qui tirera profit de l'amélioration des services d'archivage des résultats de recherche.

D'après les preuves recueillies, il est évident qu'aucun établissement national n'est mandaté pour conserver, gérer et rendre accessibles les résultats de recherche, que des besoins importants au sein de la communauté des chercheurs en matière d'archivage des données ne sont pas satisfaits et que la création d'un système national d'archivage des résultats de recherche procurerait des avantages importants. La forme que devrait prendre un tel système n'a toutefois pas encore été examinée.

Il est également évident que même si les chercheurs canadiens recueillent chaque année un important volume de résultats de recherche, les opinions en ce qui concerne le partage de ces résultats et les conditions du partage ne sont pas unanimes. Même si le milieu des archives et des bibliothèques est convaincu de la nécessité d'un service national d'archivage des données, un nombre important de chercheurs ne sont pas convaincus ou ont une connaissance limitée de la valeur des archives des résultats de recherche.

À ce stade-ci, le groupe de travail qui a procédé à l'évaluation des besoins recommande qu'on l'autorise à passer à la phase deux de la consultation et à examiner la forme la plus appropriée d'archives nationales de résultats de recherche pour le contexte canadien.

# Résultats de recherche archivés dans une société du savoir



**Une économie du savoir est basée sur la capacité de ses participants à transformer les données, les informations et les idées en nouvelles connaissances. Les archives de données, comme une nouvelle pièce d'équipement diagnostic, sont un outil innovateur.**

## 1.0 Introduction

En mai 2000, une chercheuse a demandé à l'archiviste de données de l'University of Toronto si elle pouvait lui obtenir un fichier de données statistiques intitulé *Access to Justice in Ontario, 1985-1988*. Une demande simple. Le fichier est cependant introuvable en Ontario ou ailleurs au Canada, le seul endroit où on peut le trouver, c'est à l'Inter-University Consortium for Political and Social Research (ICPSR) de l'université du Michigan. La chercheuse a sans aucun doute reçu le fichier demandé par voie de communication protégée, sans être altéré, en étant compatible avec le logiciel disponible et doté de métadonnées. Elle n'a pas eu de problème à l'obtenir, mais si elle avait fait partie d'une autre université, une qui n'est pas membre de l'ICPSR, sa demande n'aurait peut-être pas été traitée en priorité, on lui aurait peut-être demandé des frais ou on encore refusé sa demande.

Ce rapport prouve qu'il y a une lacune importante dans l'infrastructure de recherche canadienne. De plus en plus, les sociétés modernes dépendent des connaissances pour leur fonctionnement. Les connaissances sont le produit d'informations fiables, qui à leur tour, sont fondées sur des données fiables provenant d'une variété de sources sérieuses. Au Canada, il n'y a aucun organisme national ou institutionnel ni de réseau en place pour garantir la conservation et l'accessibilité des résultats de recherche aux fins de réutilisation, de vérification et de reproduction des conclusions, d'analyses comparatives ou de représentation des intérêts du Canada sur la scène internationale. Par conséquent, de précieux résultats de recherche sont perdus ou stockés ailleurs et ainsi ne sont pas intégrés au processus continu d'acquisition de connaissances. Comme l'a exprimé un des intervenants :

*« Les Canadiens, ont d'une façon ou d'une autre, payé un prix pour créer ces fichiers, parce qu'on estimait qu'il était nécessaire de recueillir certains renseignements. Toutefois, ces fichiers de données offrent plusieurs possibilités. En effet, les chercheurs peuvent s'y référer plusieurs années plus tard et découvrir de nouveaux liens. Si on perd les fichiers de données, les sommes investies pour leur création ne procurent pas le rendement qu'aurait pu donner la conservation. »*

À l'heure actuelle, il est techniquement possible pour un chercheur de Saskatoon ou de St. John de savoir non seulement quelle bibliothèque détient un livre ou une revue précise, mais aussi de savoir quelles archives contiennent les ensembles de données ayant servies à leur publication. Le chercheur peut également obtenir la publication ou les archives sur son ordinateur. Le système de livraison CA\*Net3 est en place, et divers projets de revues électroniques sont en cours. L'élément manquant, c'est un service national d'archivage des résultats de recherche. Un dépôt complet et moderne d'ensembles de résultats de recherche offre la possibilité de créer des synergies entre les bibliothèques et les archives, permettant aux chercheurs de tout le pays d'avoir accès à la fois à des documents publiés, à des documents d'archives et à des ensembles de données connexes provenant du même endroit.

Un bon nombre de projets de recherche universitaires et d'infrastructures de recherche sont en cours actuellement, et une fonction d'archivage des résultats de recherche leur sera nécessaire. Voici quelques exemples :

- Le centre de ressources visuelles de l'Université McGill, qui mettra au point un dépôt numérique d'images visuelles, d'artéfacts commerciaux, de photographies et de documents culturels;

- Le Text Analysis Portal for Research, un projet conjoint de l'University of Victoria, de l'University of Alberta, de l'University of Toronto, de McMaster University, de l'Université de Montréal et de l'University of New Brunswick qui donnera l'accès Internet à une grande variété de bases de données textuelles et d'outils logiciels;
- The Supply Chain Research Institute de l'Athabasca University, qui élaborera une série de bases de données pour soutenir la recherche sur les pratiques de gestion organisationnelle;
- La Video Archive de Brock University, qui donnera accès à des enregistrements numériques de productions télévisuelles et cinématographiques.

## Définition d'une fonction nationale d'archivage des résultats de recherche

Pour les besoins de ce rapport, la fonction nationale d'archivage des résultats de recherche signifie conserver, gérer et rendre accessibles au public des informations numériques structurées de façon méthodique en vue de produire des nouvelles connaissances. Une telle fonction permet de gérer les résultats de recherche existants entre les données brutes et les résultats de recherche publiés. Les acquisitions pourraient comprendre des informations numériques produites par les chercheurs et les informations qui les intéressent. La présente définition fait ressortir six points :

- La conservation, la gestion et l'accès sont trois des aspects importants de l'archivage.
- Les documents à être archivés sont des résultats de recherche sous forme numérique. Même si plusieurs intervenants estiment que l'objectif d'une fonction nationale d'archivage des résultats de recherche serait de donner accès à des publications imprimées et électroniques, il existe des établissements mis sur pied pour répondre à ce besoin.
- Les documents ou informations numériques recueillies de façon méthodique et dans un but précis sont un élément essentiel. Les informations ne sont pas recueillies de façon désordonnée.
- L'objectif, c'est de produire des nouvelles connaissances. Les informations numériques ne répondent pas toutes au critère de conservation des résultats de recherche.
- La fonction permettrait de relier des données brutes non traitées et des résultats de recherche publiés.
- La fonction englobe les résultats de recherche compilés par les chercheurs et les données qui les intéressent, dans les limites des ressources financières.

Tout au long de ce rapport, nous ferons référence à une « fonction » nationale d'archivage des résultats plutôt qu'à un organisme ou à une institution. Quelle forme la fonction pourrait prendre est le sujet de la deuxième phase de consultation.

Par « fonction d'archivage des résultats », nous entendons le vaste éventail de services de conservation, de gestion et d'accès offerts par de nombreuses archives de données dans d'autres pays. Ces services comprennent la conservation hors site, l'accès Internet à des ensembles de données, des protocoles de conservation, la création de métadonnées, le transport de données entre des systèmes logiciels et des plateformes matérielles, des ressources pour la création de normes internationales et une foule d'autres services. Par « national », on entend une fonction qui offrirait ces services à tous les Canadiens et Canadiennes.

Archiver des données pour assurer leur conservation et leur accès à d'autres utilisateurs signifie aussi conseiller les chercheurs sur les meilleures pratiques, l'intervention active, la migration, la régénération, les tests et de nombreuses autres activités. Les variétés infinies de format physique de stockage, l'encodage média, les formats de fichier, les systèmes informatiques et d'exploitation signifient qu'une quantité énorme de travail doit être accomplie au moment de l'acquisition des données afin qu'elles ne soient pas touchées par la désuétude et la disparition du matériel et des logiciels ayant servi à leur création.

## 1.1 Contexte

Des archives de résultats de recherche conservent, gèrent et rendent accessibles, sous forme numérique, les documents scientifiques et culturels utilisés par les chercheurs pour bâtir notre société et notre économie du savoir. Ces fonctions constituent l'essentiel de la gestion des informations et des connaissances, mais pourtant, au Canada, il n'existe aucun établissement, organisme ni réseau pour élaborer une stratégie nationale de gestion des données ni pour prendre en charge la gestion des documents de recherche numériques. C'est la conclusion tirée lors de l'Atelier de l'OCDE/CRSH sur l'infrastructure de la recherche en sciences sociales, tenu à Ottawa en octobre 1999. De même, des organismes comme l'Association canadienne des utilisateurs de données publiques et des chercheurs, le Programme canadien des changements à l'échelle du globe et des chercheurs comme M. John English, Ph.D., présentant son rapport sur les Archives nationales et la Bibliothèque nationale, ont lancé un appel en faveur de mesures pour régler la question de l'archivage et de la gestion des données au Canada.

Vers le milieu de l'an 2000, le CRSH, de concert avec les Archives nationales, a entrepris une consultation en deux étapes sur l'archivage des résultats de recherche. Dans la première étape, on a évalué le besoin parmi le milieu des chercheurs et des archivistes d'une installation nationale d'archivage des résultats de recherche. Si le CRSH et l'Archiviste national déterminent qu'il y a un besoin suffisant, le Groupe de travail passera à la deuxième étape, durant laquelle il examinera la meilleure approche et fera des recommandations sur l'organisation et les fonctions d'un organisme approprié.

Au Canada, la collecte des résultats de recherches est une énorme entreprise. À lui seul, le gouvernement fédéral y consacre plus de 1 milliard de dollars chaque année. On estime que les gouvernements provinciaux, les organismes à but non lucratif et les chercheurs universitaires dépensent au moins autant à cette fin. La somme dépensée par le secteur privé est inconnue. Selon les calculs les plus prudents, le total des ressources financières consacrées à la collecte des données se chiffre dans les milliards de dollars chaque année. Garantir un profit maximal de ces investissements est une responsabilité partagée par tous les chercheurs et les organismes de recherche financés par les fonds publics.

L'évolution rapide et l'adoption de la technologie informatique dans la recherche et les possibilités énormes de communication et de coordination offertes par Internet ont, toutes les deux, accru le besoin d'un système intégré de gestion des résultats de recherche et l'incidence possible qu'un tel système pourrait avoir sur tous les secteurs de la recherche. Cependant, l'augmentation considérable de la variété des disciplines et secteurs de recherche qui utilisent des bases de données informatiques comme outils de recherche est tout aussi importante. Les ensembles traditionnels de données statistiques ne dominent plus, et sont maintenant accompagnés de fichiers du système d'information géographique, de cartes numériques, de documents numérisés, de journaux électroniques, d'images numériques et de résultats enregistrés numériquement.

En fait, nous commençons seulement à réaliser le riche potentiel de recherche de l'information numérique dans un environnement informatisé et branché sur le Web. On peut maintenant créer des ensembles de données à partir d'informations quantitatives et qualitatives dans pratiquement tous les formats. Grâce à Internet, ces ensembles de données peuvent être transportés à n'importe quelle distance en quelques secondes. Avec le matériel et le logiciel appropriés, ils peuvent être recherchés, manipulés, transformés et intégrés facilement. Gardée en mémoire sous forme numérique, l'information ne requiert qu'une fraction de l'espace nécessaire à des documents sur papier ou sur film. Cela accroît considérablement l'efficacité et la portée éventuelle des activités de recherche.

Depuis le milieu des années 60, bon nombre de pays ont reconnu l'utilité d'archives de données unifiées et coordonnées et ont construit des installations nationales selon une vaste gamme de modèles institutionnels. Certains sont des exploitations hautement centralisées financées par l'État. D'autres ont recours à des frais d'utilisation en plus des fonds publics. La majorité se

trouve dans des universités pour mieux desservir les chercheurs. Les collections de résultats de recherche couvrent tout, des résultats de sondages et d'enquête à des œuvres littéraires, en passant par des émissions de télévision et des renseignements médicaux. Les services offerts comprennent des laboratoires informatiques sur place, la commande automatique d'ensembles de données auprès d'autres dépôts, l'élaboration et l'évaluation de logiciels, des programmes de formation, des guides et protocoles pour la création d'ensembles de données, l'élaboration de métadonnées, le transport de données à travers les technologies, du matériel didactique adapté, des réseaux Web complets et des bureaux régionaux.

Aujourd'hui, personne n'est certain du nombre d'ensembles de données en sciences humaines et sociales qui existent au pays. Personne, non plus, n'est certain du lieu où ils se trouvent, de ce qu'ils contiennent ou s'ils sont conservés adéquatement. Il n'existe pas de répertoire principal pour une discipline donnée, sans parler des champs d'études généraux. Nous savons que des ensembles de données importants et irremplaçables ont été perdus — tous les sondages Gallup d'avant 1947 et la plupart d'avant 1952, une part importante des sondages d'opinion pris avant le référendum de 1995 au Québec — mais il y a si peu de documentation que nous n'avons aucune idée de l'importance de la perte. Des preuves semblent cependant indiquer qu'elles sont importantes et qu'elles ont des répercussions énormes sur les chercheurs.

La gestion, la préservation et l'accessibilité des résultats de recherche deviennent de plus en plus importantes pour les chercheurs de nombreuses disciplines. Cela est particulièrement vrai dans les études interdisciplinaires où les chercheurs recherchent de l'information d'une vaste gamme de sources souvent peu connues. Repérer des ensembles de données à l'extérieur de sa propre discipline peut devenir une tâche frustrante et fastidieuse. Et, comme de plus en plus de recherches en sciences humaines et sociales prennent une orientation axée sur les problèmes, plutôt que sur les disciplines, la nécessité de coordonner et de rendre accessible le matériel de recherche augmente.

*Archiver des données, c'est plus que faire une copie de sauvegarde... L'archivage des données exige d'offrir à long terme les ressources, l'expertise et le service public requis pour assurer l'accès continu aux fichiers de données, décrire et documenter les fichiers et assurer l'accès à ces fichiers ainsi que leur contrôle intellectuel. Une des raisons pour lesquelles les chercheurs ne sont pas emballés par l'idée, c'est parce qu'il est difficile de savoir quelles données ont été recueillies. Il est logique de faire des économies d'échelle et de centraliser les ressources requises pour une entreprise de cette envergure.*

Réponse d'une personne interrogée

Le CRSH, les Archives nationales et à vrai dire le Canada ne sont pas les seuls à être aux prises avec des problèmes d'archivage des résultats de recherche. Un bon nombre de pays ont récemment créé des archives nationales de résultats de recherche, parmi ceux-ci, la Finlande, le Japon, l'Estonie, la Slovénie et l'Afrique du Sud. Dans certains pays, même les installations existantes d'archives de données mènent des consultations. Le conseil de recherche économique et sociale de la Grande-Bretagne (The Economic and Social Research Council) est à revoir sa politique d'archivage des données. Depuis le milieu des années 90, l'administration américaine des archives et des dossiers et le National Research Council étudient des façons de gérer et de conserver la grande quantité de données scientifiques et techniques produites en sciences naturelles grâce au financement de l'État. Il y a quelques années, le Conseil international des unions scientifiques (CIUS) a mis sur pied CODATA, un comité interdisciplinaire chargé de la gestion des données, du contrôle de la qualité et de la diffusion des données provenant de toutes les disciplines scientifiques et techniques. Depuis quelques années, le Conseil international pour l'information scientifique et technique (ICSTI) commande des rapports

d'enquête et organise des ateliers internationaux en vue de tenter d'établir une méthode d'archivage coordonnée qui englobe les limites nationales et disciplinaires.

Comme l'indique un rapport clé commandé par l'ICSTI, tous les milieux et disciplines scientifiques sont confrontés aux problèmes de gestion, de conservation et d'accessibilité des données de recherche, y compris les sciences humaines.<sup>1</sup> L'informatique facilite le stockage numérique d'une diversité croissante de résultats de recherche. Les cartes, photographies, enregistrements sonores, l'hypertexte, les pages dynamiques, les systèmes d'information géographique, le multimédia et la vidéo interactive sont largement utilisés à l'heure actuelle et exigent une gestion et une conservation efficaces si on veut continuer de les utiliser dans le processus de création des connaissances.

## 1.2 Processus de consultation

Le CRSH et les Archives nationales du Canada ont demandé à un Groupe de travail d'experts de divers domaines de mener une enquête approfondie en vue d'évaluer le besoin d'une fonction ou d'un service national d'archivage des résultats de recherche. Le Groupe de travail s'est penché sur les questions suivantes :

- Dans quelle mesure a-t-on besoin d'une fonction d'archivage de données unifiée et coordonnée? Est-ce qu'on pourrait répondre aux besoins actuels et futurs en apportant de légères modifications aux politiques et aux mécanismes existants?
- Quelles sont les lacunes dans les mandats et les structures des établissements existants pour ce qui concerne la gestion des données de recherche?
- Qui bénéficiera de l'amélioration de la gestion des données de recherche et dans quelle mesure?
- Comment l'efficacité de la gestion, la conservation et la consultation des données de recherche favorisera-t-elle la capacité de recherche du Canada?

Deux groupes de chercheurs émérites et de spécialistes en archivistique et en bibliothéconomie ont participé aux consultations. Un groupe de travail composé de neuf personnes travaillant de concert avec un groupe d'étude plus important a mené les consultations et rédigé le rapport de la phase 1. Présidé par le D<sup>r</sup> John ApSimon, ancien vice-président de la recherche à Carleton University, les groupes de travail et d'étude ont réuni des spécialistes des sciences sociales et humaines, le directeur d'un important centre de recherche en communications, un archiviste de données du milieu universitaire possédant une vaste expérience en promotion de systèmes nationaux et un représentant du Conseil du Trésor fédéral qui travaille à l'élaboration d'une infrastructure nationale dans les environnements numériques. Les deux groupes ont discuté en tête-à-tête.

La consultation visait à recueillir les commentaires d'une grande variété de sources bien informées. Le Groupe de travail a interrogé les personnes suivantes : les chercheurs universitaires et les archivistes et bibliothécaires des milieux universitaires et gouvernementaux. Ce sont les principaux producteurs, utilisateurs et gestionnaires de résultats de recherche, et on a estimé qu'ils seraient les mieux placés pour connaître les besoins en matière d'archivage des résultats de recherche au Canada.

La première phase de l'évaluation des besoins comprenait les activités suivantes :

- Envoi de lettres à 246 directeurs d'instituts ou de groupes de recherche universitaires annonçant la consultation et les invitant à faire part de leurs commentaires et opinions sur les besoins en archivage de données.

---

<sup>1</sup> *Digital Electronic Archiving: The State of the Art and the State of the Practice*, ICSTI.

- Tenue d'une assemblée publique des intervenants aux Archives nationales en octobre 2000. 55 personnes de milieux professionnels très variés ont participé aux discussions.
- Envoi de lettres aux sous-ministres de ministères fédéraux dont le mandat comprend la recherche.
- Présentations du personnel du CRSH à des groupes précis, comme au Réseau canadien pour l'avancement de la recherche, de l'industrie et de l'enseignement (CANARIE), au groupe des projets d'enseignement virtuel et aux chercheurs de l'University of Calgary, de l'University of Alberta, de l'Université McGill et de Brock University.
- Création, de concert avec Ressources naturelles Canada, d'un site Web accessible au public, qui contient des informations générales et un forum de discussion.
- Réalisation par des membres des groupes de travail et d'étude d'enquêtes portant notamment sur les statistiques sociales, les données géospatiales, l'histoire, l'analyse de texte, la sécurité et l'authentification de documents électroniques.

On a envoyé des questionnaires d'évaluation des besoins à quatre groupes de producteurs de données, d'utilisateurs de données et d'archivistes :

- 1) Un échantillon stratifié de 20 p. 100 de tous les chercheurs ayant reçu une subvention de soutien à la recherche du CRSH entre 1998 et 2000.
- 2) Un groupe d'intervenants définis pendant le processus de consultation, dont des ministères fédéraux et provinciaux, des organisations du secteur privé, des associations d'archives, des associations universitaires, des organismes qui offrent des subventions de recherche, des bureaux de recherche universitaire, des associations de bibliothèques universitaires et des chercheurs individuels.
- 3) Les directeurs des dix dépôts de données listés par le CRSH dans le *Guide des détenteurs de subventions*.
- 4) Des personnes ressources de l'Initiative de démocratisation des données travaillant dans 66 universités canadiennes. Ce sont des bibliothécaires professionnels chargés de la gestion des fichiers de données de Statistique Canada accessibles au public ainsi que d'une grande variété d'autres résultats de recherche.

Les taux de réponse variaient entre 20 p. 100 pour le groupe des intervenants et 70 p. 100 pour les directeurs de dépôts de données. Le taux de réponse des chercheurs subventionnés par le CRSH est de 26 p. 100. Les personnes qui ont répondu provenaient de toutes les régions du Canada et comprenaient des francophones et des anglophones. Les questionnaires remplis par les intervenants, les directeurs et les personnes-ressources de l'Initiative de démocratisation des données révèlent une plus grande familiarité avec les problèmes d'archivage des données, en particulier les problèmes complexes.

*Les données générées par la recherche financée par l'État, les instruments de recherche utilisés, le modèle utilisé, les cadres d'échantillonnage, etc. devraient être archivés et rendus accessibles aux autres chercheurs. Cet archivage favoriserait la collaboration et servirait aux études longitudinales, aux études de réplification, aux études comparatives, à la création de modèles de questions « normatives » et à des analyses secondaires. La transparence et la reddition de comptes seraient encouragées en exigeant l'archivage et l'accessibilité des données. En outre, la prise en compte de données de ce genre devrait devenir une caractéristique plus importante de la planification de « nouvelle » recherche primaire. Ainsi, on ne réinventerait pas la roue et il en résulterait peut-être des travaux plus créatifs. C'est du moins ce que j'en pense, pour ce que ça vaut.*

Réponse d'une personne interrogée

## 1.3 Réponses aux questionnaires

### Chercheurs subventionnés par le CRSH

On a interrogé un échantillon stratifié de 20 p. 100 de tous les chercheurs ayant reçu une subvention de soutien à la recherche du CRSH entre 1998 et 2000. On a reçu 116 réponses. Chaque chercheur a reçu un questionnaire portant sur son expérience de la création et l'archivage de résultats de recherche ou de l'utilisation de résultats de recherche produits par d'autres. On leur a également demandé ce qu'ils pensaient des questions fondamentales à la base des principes de l'archivage des données et d'évaluer l'importance de la mise sur pied des services nationaux de conservation des résultats de recherche.

Pour une année donnée, la moitié des chercheurs ayant reçu une subvention de soutien à la recherche du CRSH ont produit des résultats de recherche. Pour ceux qui ont répondu au questionnaire, le nombre atteint 55 p. 100. Ce chiffre dépasse les quelque 1 200 ensembles de données créés par les chercheurs ayant reçu une subvention de soutien à la recherche du CRSH entre 1998 et 2000, ou une moyenne de 400 par année. Depuis janvier 2001, seulement 7 p. 100 des chercheurs interrogés ont archivé leurs données et 18 p. 100 indiquent qu'ils ont l'intention de le faire. Sur les 18 p. 100 qui ont l'intention d'archiver leurs données, moins de la moitié ont pu indiquer un service ou un organisme d'archivage des données. Même selon le meilleur scénario, cela signifie que sur une période de trois ans, nous serons privés de près de 950 ensembles de données financés par l'État.

Que pensent les chercheurs des problèmes de cueillette et d'archivage des données?

- 81 p. 100 indiquent que les données qu'elles soient créées par eux ou par d'autres sont un précieux sous-produit de la recherche.
- 79 p. 100 croient que les résultats de recherche sont la propriété intellectuelle du chercheur principal.
- 78 p. 100 mentionnent que l'analyse secondaire des données existantes est une méthodologie de recherche valable.
- 73 p. 100 estiment que les données ne devraient être partagées que si le chercheur principal y consent.
- 71 p. 100 indiquent que les conseils de recherche devraient assumer les coûts de la préparation des données pour leur partage.
- 68 p. 100 conviennent que les chercheurs sont responsables de l'administration des données qu'on ne peut pas facilement reproduire.
- 50 p. 100 estiment qu'affecter des fonds à la préparation des données pour l'archivage ne constitue pas du gaspillage.
- 48 p. 100 conviennent que l'archivage devrait faire partie intégrante de la recherche.

Voir l'Annexe 1 pour une analyse complète du questionnaire.

### Dépôts de données listés par le CRSH

En 1990, le CRSH a publié une nouvelle annexe à son guide à l'intention des candidats contenant une liste de 10 informathèques universitaires prêtes à servir de dépôts pour les résultats de recherche issus des recherches financées par le CRSH. Depuis la publication de cette annexe, le CRSH a accordé plus de 8 500 subventions, mais moins de dix de ces projets ont déposé leurs données. On a entrepris l'étude de ces dix informathèques pour comprendre l'efficacité de la méthode existante d'archivage et le niveau de soutien que ces informathèques sont capables d'offrir pour l'archivage des résultats de recherche. Sept des dix dépôts ont rempli le questionnaire.

Lorsqu'on a demandé aux informathèques si elles avaient reçu des fichiers de projets financés par le CRSH depuis la publication du répertoire de dépôts, trois des sept informathèques ont répondu non. Sur les quatre dépôts de données ayant reçu des fichiers, un avait reçu des fichiers de trois projets, deux avaient reçu des fichiers de deux projets et un avait reçu des fichiers d'un projet. Lorsqu'on leur a demandé d'évaluer le succès de la liste des dépositaires de données du *Guide des subventions du CRSH* pour inciter les chercheurs à déposer leurs résultats, cinq ont indiqué que c'était un échec complet et deux, un échec. Suivant cette question, on leur a demandé ce qu'ils pensaient de l'efficacité de la liste de dépôts à remplir la fonction d'archives de données au Canada. Six ont répondu que la liste de dépôts était vraiment inefficace et une qu'elle était inefficace.

Lorsqu'on leur a demandé quels facteurs déterminaient si un chercheur allait mettre ses données en dépôt, ils ont répondu que tout dépendait des pratiques actuelles des chercheurs en matière de données et des responsabilités des conseils qui accordent des subventions en matière de soutien à la conservation des données. Premièrement, les personnes qui ont répondu estiment que les chercheurs ont peu de connaissances en archivage de données ainsi que des exigences du CRSH.

Deuxièmement, ils ont mis l'accent sur les responsabilités des conseils de recherche. La politique actuelle du CRSH sur le dépôt de données est perçue comme n'étant pas suffisamment sévère ni exécutoire. Une personne a mentionné la nécessité pour le Conseil de faire un suivi pour évaluer la conformité avec la politique en matière de dépôt. Deux ont proposé des sanctions que pourrait prendre le Conseil dans l'éventualité où les chercheurs ne mettraient pas leurs données en dépôt. Trois ont indiqué que la préparation des données pour leur dépôt n'est pas couverte par les conseils qui accordent des subventions. Un des intervenants résume son point de vue :

*« [Les chercheurs] auront besoin d'instructions et d'aide pour la création de documentation adéquate de leurs données afin que les autres puissent facilement les utiliser. Nous avons reçu deux ensembles de données non accompagnées de documentation. Les chercheurs principaux n'avaient pas documenté leur étude. »*

## **Intervenants**

Le questionnaire destiné aux intervenants était axé sur trois questions précises : estiment-ils important de créer une fonction nationale d'archivage des données, qui pourrait s'en servir et savent-ils si des données sont actuellement susceptibles d'être perdues ou détruites. Quarante-trois personnes ont rempli le questionnaire, ce qui représente un taux de réponse de 20 p. 100. Quatre-vingt-huit pour cent de celles-ci proviennent du milieu universitaire et 12 p. 100 du gouvernement fédéral ou provincial.

La première question du questionnaire portait sur l'importance du soutien national pour la conservation des résultats de recherche. La grande majorité des intervenants (95 p. 100) estiment qu'il est « très important » (70 p. 100) ou « important » (25 p. 100) que le Canada mette sur pied un service d'archivage. Tous les intervenants du secteur gouvernemental estiment que le soutien national pour la conservation des résultats de recherche est « très important », tandis que 68 p. 100 de ceux du milieu universitaire sont du même avis (les 27 p. 100 restant jugent important le soutien national).

Soixante-quatre pour cent des personnes ayant répondu indiquent qu'elles sont au courant de cas où des données pourraient être perdues. Dans une autre question, dont le taux de réponse est de 27 p. 100, on leur a demandé de décrire les données qu'elles croyaient vulnérables. La plupart des descriptions consistaient en des généralisations au sujet des types de données plutôt qu'en des titres de collections de données précises. Deux catégories générales se sont

dégagées des réponses données. Dans la première, on relève des données provenant d'une variété de chercheurs. Elles comprennent des données issues de recherches financées par le CRSH, de producteurs de données géospatiales, de thèses ou de mémoires de recherche d'étudiants diplômés, d'unités de recherche de ministères fédéraux et provinciaux, de bases de données communautaires et régionales, de projets de petite envergure ainsi que de sondages privés et financés par l'État. La deuxième catégorie consiste en des raisons techniques pour lesquelles les données sont en péril. Par exemple, des données stockées sur support traditionnel et des données qui n'ont pas été documentées de façon adéquate.

Les intervenants ont nommé plusieurs projets vulnérables, dont le sondage canadien sur les maladies, le sondage national de Nutrition Canada, le sondage ontarien sur la santé, le projet d'expédition du Canada atlantique et le projet sur les familles canadiennes.

### **Personnes-ressources de l'Initiative de démocratisation des données**

Les personnes-ressources de l'Initiative de démocratisation des données consistent en un sous-groupe de spécialistes du milieu universitaire chargés d'assurer l'accès aux documents de Statistiques Canada et à d'autres résultats de recherche détenus dans leur établissement. Par conséquent, ce groupe connaît l'importance des données pour la recherche. Il travaille à donner l'accès aux résultats de recherche au nom des chercheurs et des étudiants et connaît les problèmes d'accessibilité aux données. On a reçu 31 réponses de ce groupe à un questionnaire portant essentiellement sur l'importance des archives de résultats de recherche pour divers secteurs et activités de recherche au Canada. En outre, on leur a posé des questions précises sur l'efficacité des méthodes existantes à conserver les données et on leur a demandé de nous dire à qui il incombe de conserver les résultats de recherche.

Les bibliothécaires et le personnel assurant l'accès aux données dans leur établissement reconnaissent l'importance de conserver les résultats de recherche universitaire ainsi que de tous les paliers de gouvernement. Ils trouvent moins important de conserver les résultats de recherche du secteur privé. C'est peut-être parce qu'ils ne trouvent pas que c'est un objectif réaliste. La plupart s'entendent pour dire que les bibliothèques ne sont pas en mesure d'offrir des services d'archivage de données. Tous reconnaissent cependant la nécessité de services nationaux d'archivage des résultats de recherche.

### **1.4 Présentations des intervenants**

Les vingt présentations faites par les bibliothécaires-archivistes et les chercheurs universitaires et un échantillon d'opinions recueillies d'un bout à l'autre du pays se recoupent : tous sont en faveur de la création d'archives nationales de résultats de recherche (voir Annexe 3). Les facultés représentées, dont les recherches sont financées par le CRSH et d'autres organismes de financement provinciaux et fédéraux, étaient les humanités, les sciences de l'information et les sciences sociales. Elles tirent toutes profit des résultats de la recherche et y contribuent. Les exposés soulignent que seule une fonction nationale d'archivage des résultats de recherche peut garantir la crédibilité et la créativité des recherches fondées sur les résultats.

Les observations et commentaires des intervenants ainsi que de ceux ayant répondu aux questionnaires et ceux des membres des groupes de travail et d'étude sont présentés ci-dessous.

## 2.0 Lacunes dans l'infrastructure de recherche existante

En ce qui concerne la conservation, la gestion et l'accessibilité des documents de recherche numériques, des lacunes importantes existent dans les mandats et structures des institutions et organismes fédéraux et des bibliothèques et archives universitaires. Les archivistes de divers secteurs — fédéral, provincial, postsecondaire et public, indiquent qu'ils ne peuvent pas résoudre séparément ou conjointement le problème historique de gestion des données sans la présence d'un organisme ou d'une fonction nationale d'importance d'archivage des résultats de recherche. À l'heure actuelle, de nombreuses archives gèrent les résultats de recherche, mais aucune ne s'estime capable d'assurer l'infrastructure centrale essentielle.

### 2.1 Mandats des Archives nationales du Canada et de la Bibliothèque nationale du Canada

Jusqu'en 1986, date de sa fermeture, la division des archives ordinolingues des Archives nationales du Canada acceptait les documents de recherche numériques. Aujourd'hui, les efforts de préservation des Archives nationales sont axés sur les documents du gouvernement fédéral et les documents privés d'importance nationale, qui sont définis comme des documents qui « témoignent des efforts et des expériences des personnes, groupes, institutions, personnes morales et autres organisations qui ont obtenu une reconnaissance sur le plan national ou international. Elles documentent également l'environnement physique du Canada ainsi que des événements et tendances (culturelles, politiques, économiques, sociales, démographiques, scientifiques et religieuses) ayant une portée nationale »<sup>2</sup> [trad.]

La stratégie d'acquisition de 1994 des Archives nationales définit dix thèmes qui servent de guide dans le choix des acquisitions. Cette approche thématique vise à créer un vaste échantillon représentatif renfermant des informations importantes ou uniques qui enrichiront considérablement notre compréhension de l'histoire, de la société, de la culture et des habitants du Canada, comme le recensement national de Statistique Canada. Bien qu'il s'agisse d'une excellente approche pour les Archives nationales, compte tenu de son mandat général et de ses ressources limitées, elle ne répond que partiellement aux besoins d'archivage des milieux de recherche universitaires et autres. Les chercheurs doivent avoir accès à plus d'un échantillon représentatif.

Au bout du compte, les Archives nationales sont une institution fédérale responsable de préserver la mémoire de la nation. En revanche, le milieu canadien de la recherche exploite un cadre analytique géographique, social, culturel et économique beaucoup plus vaste. Les chercheurs créent des données afin de faciliter l'examen de réalités sociales, économiques et culturelles dépassant souvent les frontières du Canada. Le mandat des Archives nationales n'englobe tout simplement pas les besoins variés des chercheurs universitaires canadiens en matière d'archivage des résultats de recherche.

C'est également le cas de la Bibliothèque nationale du Canada. Son mandat est de préserver le patrimoine canadien de l'édition. Au cours des dernières années, son mandat s'est élargi, on y a ajouté des publications et des revues électroniques ainsi que des sites Web, mais les résultats de recherche sont pratiquement exclus. Les ensembles de données ne sont pas publiés et par le fait même exclus du mandat de la Bibliothèque. Qui plus est, à l'heure actuelle, ni la Bibliothèque ni les Archives nationales ne possèdent l'expertise statistique et technique nécessaire pour combler cette lacune de l'infrastructure de recherche canadienne.

---

<sup>2</sup> Archives nationales du Canada, *An Overview of the Acquisition Policy of the National Archives of Canada*.

## **2.2 Politique du Conseil de recherches en sciences humaines du Canada en matière d'archivage des données**

En 1990, le Conseil de recherches en sciences humaines du Canada (CRSH) a adopté une politique précise relativement aux ensembles de données produits par les chercheurs ayant recours à des fonds publics. Tel qu'indiqué dans la politique relative aux demandes :

« Le CRSH exige que les données recueillies avec son aide, y compris les fichiers exploitables par des machines ou les bases de données informatisées, deviennent propriété publique et soient mises à la disposition des autres chercheurs dans un délai raisonnable, à condition que le caractère confidentiel de certains renseignements et le droit à la protection des renseignements personnels soient respectés. En conséquence, le Conseil exige également que l'établissement auquel le chercheur principal est rattaché ou tout autre établissement qui devient dépositaire de données prenne les mesures nécessaires pour conserver les données et en faciliter l'accès aux chercheurs. »

La politique donne une liste des universités où les chercheurs peuvent y déposer leurs ensembles de données. On y indique également que les coûts associés à la préparation des données pour l'archivage font partie des dépenses admissibles.

Cette politique vise à favoriser le développement des connaissances en sciences sociales et humaines en encourageant le partage des résultats des recherches entre les chercheurs. Le partage des données renforce notre capacité collective à répondre aux normes universitaires de transparence en permettant d'approfondir l'analyse, la réplication, la vérification et l'amélioration des résultats de recherche. Ces possibilités renforcent le développement des champs de recherche et le potentiel de travail interdisciplinaire. En outre, une meilleure disponibilité des résultats de recherche peut contribuer à améliorer la formation des étudiants de premier et deuxième cycles et rendre possible d'importantes économies d'échelle grâce à l'analyse secondaire de données existantes. Enfin, les chercheurs dont les travaux sont financés par l'État ont l'obligation de faire preuve de transparence et de rendre des comptes.

Malheureusement, cette politique n'a pas atteint ses objectifs. Lorsqu'on a demandé aux chercheurs subventionnés par le CRSH s'ils avaient créé des ensembles de recherche au cours de leurs derniers projets de recherche, 80 p. 100 de ceux qui avaient répondu oui n'avaient pas archivé leurs données. En réalité, sur une période de 11 ans, seulement 10 ensembles de données ont été déposés chez un dépositaire listé dans le guide du CRSH.

Parmi les chercheurs sondés qui ont archivé des données issues de projets financés par le CRSH ou qui avaient l'intention de le faire, plusieurs ne savaient pas où envoyer leurs ensembles de données. D'autres croyaient que les bibliothèques universitaires s'occuperaient de les entreposer. Un certain nombre d'autres chercheurs avaient l'intention de recourir aux archives américaines ou à des institutions canadiennes non dotées de service d'archivage. Et un chercheur a indiqué : « nous archiverons nous-mêmes nos données puisque nous ne sommes pas disposés à les archiver dans des archives étrangères. »

## **2.3 Services universitaires de traitement des données**

Les archivistes et bibliothécaires universitaires indiquent que les universités n'ont pas les ressources nécessaires pour conserver, modifier et mettre à jour des bases de données de plus en plus importantes et complexes et qu'elles ne sont pas en mesure d'offrir le vaste éventail de services dont les chercheurs ont besoin. Leur principale préoccupation consiste à offrir à leur clientèle locale l'accès à des fichiers de données facilement accessibles. Ces établissements

sont confrontés aux mêmes problèmes associés à la gestion, à l'accessibilité et à la conservation des résultats de recherche tout en essayant quotidiennement de combler les lacunes résultant de l'absence d'un organisme national doté de suffisamment de ressources, d'expertise et de pouvoir pour offrir les services nécessaires.

Les archives canadiennes existantes de résultats de recherche universitaire n'ont pas suffisamment de personnel ni d'équipement. Sur les sept dépositaires de données qui ont répondu au questionnaire, aucun ne possède plus de 2,5 employés à plein temps et un a moins d'un employé à plein temps. Ces chiffres soulignent leur incapacité à gérer le vaste éventail de services nécessaires à l'archivage efficace des résultats de recherche.

## **2.4 Authentification des résultats de recherche**

L'authentification est un des moyens utilisés pour vérifier l'identité et l'intégrité des données. Elle fait référence à la validation des fichiers de données et/ou de la personne qui accède aux données à un moment précis. L'authentification assure que les personnes qui accèdent aux fichiers de données sont des utilisateurs valables et qu'ils n'altéreront pas les données. La présence de cette authentification implique que, au moment de leur utilisation, les données en question ont été vérifiées pour s'assurer de leur authenticité et créées par la ou les personnes réputées les avoir créées. Ce n'est toutefois pas suffisant pour assurer l'authenticité permanente.

Les procédures d'authentification intégrées aux processus de création, de transmission, de réception, d'utilisation, de mise à jour et de conservation des fichiers de données sont le moyen le plus efficace d'assurer l'authenticité des données dans le temps, surtout quand on tient compte de la nécessité de reproduire continuellement des données. Compte tenu de la désuétude rapide des systèmes électroniques, nous ne pouvons pas conserver les données en tant que tel; nous pouvons seulement préserver leur reproduction. C'est pourquoi nous avons besoin de normes nationales pour l'authentification des reproductions ainsi que la protection de leur identité et de leur intégrité. La conservation des données est une démarche préventive qui exige d'appliquer des mesures continues aux données. Seul un organisme permanent de services peut s'en charger.

## **2.5 Présence canadienne sur la scène internationale**

Le Canada a une capacité limitée pour négocier des accords internationaux d'échange de données, parce que contrairement à d'autres pays, particulièrement les pays d'Europe, il ne possède pas d'organisme national. En raison de l'absence de fonction nationale d'archivage des résultats de recherche, le Canada n'a pas voix au chapitre dans la création et l'établissement de normes internationales de résultats de recherche, de modèles de métadonnées comme l'Initiative de démocratisation des données, d'outils d'accès aux données comme le projet européen NESSTAR et le projet LIMBER ainsi que dans les projets concertés d'infrastructure internationale, comme les cadres de l'Union européenne. En outre, la représentation du Canada au sein d'organismes aussi importants que l'Association internationale des organisations de services en données et le Council of European Social Science Data Archives fait défaut.

Le partage des données est devenu essentiel à la participation à l'économie mondiale, parce que c'est grâce au partage des données que les pays peuvent obtenir des informations vitales à l'élaboration de politiques sociales et économiques adaptées au contexte mondial actuel. Même si le Canada participe à certains projets et fait partie de certaines organisations internationales, sa collaboration est, au mieux, ponctuelle. Nous ne savons même pas quelles données canadiennes sont disponibles, où elles se trouvent ou si elles sont utilisables.

### 3.0 Besoins d'archivage du milieu de la recherche

Les intervenants ont souligné qu'une fonction nationale d'archivage des résultats de recherche pourrait répondre à trois besoins d'archivage du milieu universitaire : (1) l'archivage des données que les chercheurs recueillent, mais qu'ils ne peuvent pas mettre à jour faute d'expertise; (2) l'archivage des documents informatisés du gouvernement et des établissements parapublics et privés intéressant les chercheurs et (3) l'archivage de documents informatisés de personnes qui peuvent contribuer au processus de création des connaissances ou dont les ensembles de données intéressent les chercheurs.

#### 3.1 Données perdues

Les intervenants se sont dits consternés par le fait que même si nous ne pouvons pas et ne devons pas conserver tous les résultats de recherche, un grand nombre d'ensemble de données, jugés essentiels, sont actuellement détruits. À l'heure actuelle, il est nécessaire d'améliorer l'archivage de catégories entières de résultats de recherche, comme :

- les données géospatiales
- les ensembles de données en sciences sociales et humaines créées grâce au financement du CRSH
- les ensembles de données en sciences sociales et humaines financées par d'autres sources
- les données non publiées produites par le gouvernement fédéral
- les données non publiées produites par les gouvernements provinciaux et municipaux et le secteur privé
- les données des sondages
- les données de sondages effectués par le gouvernement ou des firmes privées.

Un exemple de données perdues, ce sont les résultats de recherche dispendieuse ayant servi pour le livre de Michael B. Katz intitulé *The People of Hamilton* (Cambridge: Harvard University Press, 1975), conservées par l'Institut for Social Research de l'Université York. L'équipement ne peut plus lire les bandes magnétiques contenant les données.

#### 3.2 Occasions perdues

Un des gros problèmes auquel font face les chercheurs à l'heure actuelle, c'est la difficulté de trouver les données pertinentes à leur recherche. Il n'y pas de « répertoire collectif » d'ensembles de données conservés par les producteurs de données, les distributeurs ou les autres chercheurs. Ce qui signifie que les chercheurs doivent peut-être refaire inutilement des études coûteuses ou se fier à des données empiriques plutôt qu'à des données objectives. Une fonction nationale d'archivage des résultats de recherche pourrait éventuellement permettre aux chercheurs d'avoir accès à des sources de données depuis leur ordinateur de bureau, faisant ainsi économiser du temps, de l'argent et d'autres ressources précieuses.

Tant les chercheurs que les archivistes de données n'ont cessé de répéter que la conservation d'ensembles de données spatiales est inadéquate au Canada et que par conséquent nous perdons un grand nombre de données. Peu d'organisations produisant des données spatiales ont un service d'archivage, et souvent les vieux ensembles de données sont écrasés ou supprimés. Par ailleurs, la mise à jour continue des bases de données exige qu'on prenne des dispositions pour archiver les données à un moment précis.

### 3.3 Connaissances et attitudes des chercheurs

Au sein du milieu canadien de la recherche, nous remarquons des différences importantes dans les attitudes à l'égard de l'archivage des données ainsi que de la confusion et un manque de compréhension au sujet de l'importance et de la nature de l'archivage des données. Seulement 60 p. 100 des chercheurs subventionnés par le CRSH qu'on a interrogé comprennent et apprécient l'importance de créer un service ou une fonction nationale d'archivage des données au Canada. Une minorité importante de chercheurs ne connaissent pas ou ne sont pas certains des enjeux qui sous-tendent la conservation des résultats de recherche. Un des rôles des archives pourrait être d'informer le milieu de la recherche de la valeur des services d'archivages des résultats de recherche.

*Les chercheurs semblent peu sensibilisés à la question de l'archivage des données. Ce n'est pas parce qu'ils ne comprennent pas le concept, c'est plutôt parce que l'archivage des données ne fait pas partie de leur façon de faire. Je crois qu'il faut blâmer la formation des chercheurs et les chercheurs chevronnés qui devraient faire comprendre l'importance de l'archivage aux chercheurs débutants. Si on ne met pas l'accent sur l'importance de l'archivage dans la méthode de recherche, on ne le percevra pas comme une activité de recherche importante. Cette question doit être examinée par les associations professionnelles et les doyens de recherche chargés de la formation de la prochaine génération de chercheurs.*

Réponse d'une personne interrogée

Plusieurs archivistes universitaires ont pointé du doigt l'intérêt personnel des chercheurs et une culture de recherche qui n'est pas axée sur le partage des données. Un archiviste a mentionné « la crainte de perdre le contrôle du potentiel de recherche de leurs données — leur peur de se faire damer le pion par d'autres chercheurs qui utiliseraient leurs données déposées ». Une partie du problème, c'est que les chercheurs qui ne recueillent et analysent que leurs propres données ont le sentiment qu'elles leur appartiennent et sont réticents à les partager. En revanche, les chercheurs qui utilisent les données recueillies par d'autres organismes ou chercheurs sont beaucoup moins attachés aux données et plus conscients du besoin d'un service national d'archivage des données. En réalité, 76 p. 100 des chercheurs subventionnés par le CRSH qu'on a interrogé ont indiqué avoir utilisé des ensembles de données dans des recherches précédentes. Sur les 45 chercheurs qui ont répondu et qui ont indiqué avoir utilisé ou tenté d'utiliser des données produites par d'autres chercheurs, trois se sont vu refuser l'accès aux données. Neuf ont déclaré avoir refusé que d'autres chercheurs aient accès à leurs données, malgré le fait que cette façon de faire aille à l'encontre de la politique du CRSH.

### 3.4 Sécurité des résultats de recherche

La sécurité des données est essentielle dans le milieu de la recherche. On doit formuler et articuler les normes de sécurité au niveau national afin d'assurer la conformité dans la gestion des archives de données. Ces normes doivent porter sur : (1) les méthodes d'identification des ensembles de données et des procédures de gestion des risques pour évaluer la vulnérabilité des ensembles de données; (2) l'identification des exigences légales, réglementaires et contractuelles à prendre en compte, y compris les directives en matière d'éthique et les droits de propriété intellectuelle et (3) un ensemble de principes, méthodes et procédures que les organisations doivent suivre pour assurer la crédibilité de la création, la mise à jour sécuritaire, l'utilisation confidentielle et la conservation authentique de leurs données.

On doit élaborer des politiques et procédures de sécurité au niveau national et on doit encourager les organisations à les appliquer activement, de façon cohérente, à tous les fichiers de données qui ont une valeur nationale. Il est essentiel de procéder ainsi non seulement pour garantir que l'accès aux données est réservé aux seules personnes autorisées et que les données sont identifiables et authentiques à long terme, mais surtout pour garantir que les chercheurs et d'autres utilisateurs puissent vérifier l'intégrité et l'exactitude des données au fil du temps. Si les chercheurs ne peuvent se fier à l'intégrité des données sur lesquelles ils fondent leur recherche et si les utilisateurs ne peuvent en vérifier l'exactitude, les résultats de recherche sont inutiles.

## **4.0 Avantages d'un service national d'archivage des données**

Au bout du compte, la population canadienne bénéficiera le plus du rendement accru des investissements en recherche et de la production de plus en plus grande de connaissances rendue possible par l'utilisation répétée des ensembles de données existants. La capacité des chercheurs canadiens à étudier les changements des phénomènes sociaux, culturels, économiques et médicaux au fil du temps sera accrue par la récupération de données historiques clés ainsi que par l'accès amélioré et coordonné aux données contemporaines d'origines diverses.

### **4.1 Personnes qui en tirent profit**

Parmi les intervenants interrogés, 88 p. 100 estiment que la recherche universitaire est le plus important bénéficiaire de l'amélioration de la gestion des résultats de recherche. Les avantages comprendraient une réduction importante des coûts de recherche grâce à la non-reproduction des collections de données, à l'accès amélioré à une plus grande variété de données, qui en échange faciliteront une meilleure analyse et une plus grande interdisciplinarité, à la capacité de vérifier les résultats de recherche sans devoir reproduire les résultats de recherche et à l'accès amélioré à des ensembles de données chronologiques uniques.

En ordre d'importance, les intervenants estiment que les plus grands utilisateurs d'un système national d'archivage des données seraient (1) les chercheurs universitaires, (2) les professeurs et étudiants du niveau secondaire et postsecondaire, (3) les analystes politiques et décideurs de divers paliers de gouvernement, (4) les chercheurs du secteur privé et les organisations non gouvernementales et (5) le grand public. Plusieurs intervenants ont cependant indiqué que cette échelle est fondée sur la capacité actuelle à entreprendre des recherches à l'aide de résultats de recherche et non sur des modèles possibles d'utilisation dans l'avenir.

Le succès de la conservation et de l'accessibilité des données ne consistera pas en une méthode corrective (comme c'est le cas avec les ressources sur papier), mais plutôt en une démarche proactive où la conservation sera une caractéristique de la création des ressources. Les bibliothèques et archives universitaires, du fait qu'elles travaillent en étroite collaboration avec le milieu de la recherche qui crée ces ressources, peuvent contribuer à inculquer une culture d'archivage et de conservation efficace qui facilite l'accès continu. Au fur et à mesure que les modes/ressources de recherche évoluent (p. ex. les imposantes bases de données génétiques, les simulations dynamiques 3D, les environnements multimédia, les mondes virtuels), les bibliothèques et archives sont en mesure de montrer aux utilisateurs comment assurer la préservation et l'accessibilité des données dès le début du projet.

Les intervenants n'ont cessé de répéter que nous pouvons accroître grandement la capacité canadienne d'analyse des résultats de recherche en créant une collection cohérente de résultats de recherche avec laquelle nous pouvons ensuite élaborer des nouvelles techniques de recherche et statistiques adaptées aux conditions uniques du Canada.

## **4.2 Création de ressources pour l'analyse multidisciplinaire**

La conservation des données permet des utilisations de données novatrices et souvent imprévues entre les disciplines. Examinons un exemple précis : faire la généalogie des familles canadiennes françaises de l'époque de la Nouvelle-France au 20<sup>e</sup> siècle. Le projet a bénéficié d'importantes ressources financières du secteur public et a généré un certain nombre d'analyses historiques qui n'auraient pas pu voir le jour sans l'aide de l'informatique. Mais aujourd'hui, les nouvelles informations historiques créées par les démographes et les historiens ont d'autres utilités parce qu'elles sont disponibles en format électronique et sont documentées conformément à des règles précises. Comme les données sont disponibles sur un serveur, les chercheurs peuvent y avoir accès peu importe où ils se trouvent. L'exemple le plus frappant d'utilisation de ces nouvelles données historiques, ce sont les travaux de génétique et d'épidémiologie effectués auprès des familles québécoises à l'aide de la base de données BALSAC. Cette base de données a été conçue, mise au point et validée au cours des trente dernières années par des équipes interdisciplinaires de chercheurs en sciences humaines et sociales travaillant ensemble dans un cadre qui s'appelle maintenant l'Institut interuniversitaire de recherche sur les populations (IREP).

De nombreux chercheurs des sciences sociales — historiens, géographes, démographes, sociologues, anthropologues et juristes — et des sciences de la santé — généticiens et épidémiologistes — ont contribué à l'enrichissement de la base de données BALSAC. Dans la plupart des cas, ils ont travaillé dans des équipes interdisciplinaires, partageant leurs problèmes, méthodes et conclusions. Certains chercheurs ont relié les données de BALSAC à leurs propres données nominales, ils ont ainsi créé des bases de données multisources et pu entreprendre des recherches novatrices. En sciences sociales et humaines, la recherche a couvert des thèmes comme l'histoire et le développement de la fertilité, les migrations de population, la mobilité sociale, les clivages sociaux, l'occupation de l'espace, la transformation du monde rural, l'histoire du travail et l'histoire urbaine.

L'IREP a rendu accessible la base de données BALSAC au milieu québécois de la recherche, mais il n'y a aucune installation pour assurer sa conservation à long terme ni pour faire connaître son existence au reste du Canada. Tout aussi important, il n'y a aucun dépositaire central pour les nombreuses bases de données dérivées des travaux originaux.

## **4.3 Contribution à l'enseignement et à la formation en recherche**

Jour après jour, les professeurs d'université sont confrontés aux problèmes des étudiants qui veulent réaliser divers projets de recherche à l'aide de données. Pour une thèse de recherche de deuxième ou de premier cycle sur la politique environnementale du Canada, la capacité de trouver rapidement des ensembles de données pertinents et de définir des questions d'intérêt permet à l'étudiant de consacrer la plus grande partie de son temps à l'analyse des données. À la place, à l'heure actuelle, il arrive trop souvent que les étudiants n'aient pas assez de temps pour analyser les données à cause du temps perdu à tenter de trouver si des données existent, et si c'est le cas, à tenter d'y accéder.

Plusieurs intervenants ont observé que les étudiants en sciences sociales consacrent moins de temps à l'analyse des données qu'ils le devraient et sont nettement désavantagés lorsqu'ils poursuivent des études au deuxième cycle ou entrent sur le marché du travail comparativement aux étudiants américains qui ont facilement accès au ICPSR. De nombreux chercheurs en sciences sociales soutiennent que la capacité d'analyse de nos étudiants de deuxième cycle est inférieure à celle de leurs homologues américains. Le manque de coordination des efforts à rendre les données accessibles en est en partie responsable. Une fonction nationale d'archivage des données pourrait régler le problème.

#### **4.4 Mise en valeur des ressources en archivage des données**

Au Canada, nous manquons d'expertise dans la gestion des données et la préparation et la conservation des métadonnées, système de codage utilisé pour décrire et localiser les ensembles de données. Une fonction nationale d'archivage des données pourrait assurer la formation requise pour combler nos lacunes en gestion des fichiers de données. Les avantages potentiels s'étendent au-delà du milieu de la recherche puisque tous les secteurs de la société, privé et public, font face à de graves problèmes de gestion et de conservation des données.

En outre, à l'heure actuelle, les producteurs de données sont peu motivés à fournir des fichiers de données bien documentés puisqu'il n'y a pas de dépositaire national pour leur travail. Une fonction nationale d'archivage des données assurerait la formation des universitaires et des autres producteurs de données en création de métadonnées et en gestion des ensembles de données. Ces compétences seraient utiles au pays à l'ère de la nouvelle économie.

#### **4.5 Préservation et accessibilité des ressources textuelles électroniques**

Les ouvrages imprimés, comme les livres, journaux et manuscrits sont les principaux moyens par lesquels nous transmettons, étudions et stockons pour une utilisation future les travaux en sciences humaines et certains travaux en sciences sociales. Les philosophes, historiens, critiques littéraires, historiens de l'art, politologues et autres spécialistes des sciences sociales et humaines utilisent des sources primaires qui sont elles-mêmes des textes et produisent des nouveaux travaux de recherche qui sont également des textes. Un nombre grandissant de ces textes sont créés par ordinateur et leur format original est donc électronique. De plus, un nombre important d'universitaires canadiens ont créé des ressources de recherche textuelle « étiquetées » ou « marquées qui ne peuvent être étudiées que sous format électronique à l'aide des outils appropriés. Il y a actuellement un volume important de ressources de recherche disponibles sous forme électronique, et dans certains cas, il s'agit de l'unique format disponible. Il est prudent de dire qu'un nombre important de chercheurs doivent maintenant avoir accès à des services de texte électronique mis à jour régulièrement pour effectuer des recherches et que, dans un avenir rapproché, la majorité d'entre eux auront besoin que l'accessibilité à des procédés informatiques et à des services textuels devienne pratique courante dans la plupart des disciplines.

La Bibliothèque nationale du Canada joue un rôle essentiel dans la conservation de textes publiés, dont ceux sous format électronique, mais elle n'a pas un service d'archivage ni le mandat ou la capacité de faire face à un grand nombre des problèmes techniques complexes soulevés par le traitement des bases de données textuelles. Un des problèmes les plus importants concerne la gestion et la conservation d'outils logiciels d'avant-garde conçus pour l'analyse textuelle. Ces outils se présentent séparément ainsi qu'intégrés à des manuscrits électroniques. Des connaissances précises sont nécessaires pour assurer la compatibilité avec les nouveaux systèmes matériels et la mise à jour des liens avec d'autres bases de données textuelles ainsi que pour conseiller les chercheurs sur la meilleure façon d'utiliser ces outils.

#### **4.6 Données ne pouvant être reproduites**

Les données d'enquête statistique donnent un aperçu d'un moment précis que les chercheurs peuvent utiliser par la suite comme point de référence. En raison de leur nature, des ensembles de données de la sorte ne peuvent être reproduits. S'ils ne sont pas conservés adéquatement, on ne pourra pas s'en servir pour des recherches futures. Une fonction nationale d'archivage des données assurerait la disponibilité des données recueillies pour suivre nos progrès par rapport à un nombre d'indicateurs sociaux aux fins d'évaluation et d'examen futurs. Des données du genre nous permettraient d'évaluer objectivement les ramifications des politiques et nous donneraient

un aperçu des domaines qu'il faudrait améliorer. De cette façon, nous pourrions fonder nos programmes sociaux sur des preuves, ce qui nous permettrait de faire une meilleure utilisation de notre trésor national le plus important : nos ressources humaines.

Les grands ensembles de données d'enquête statistique, comme le recensement national, peuvent être utilisés dans des niveaux multiples d'analyse, bien au-delà de ce que Statistique Canada considère important ou est capable d'entreprendre. La conservation adéquate et l'accessibilité de tels ensembles de données permettraient aux chercheurs d'apporter un nouvel éclairage à l'analyse des données à mesure que les conditions sociales et économiques changent. À l'heure actuelle, il est difficile d'examiner les répercussions des changements sociaux. La plupart des enquêtes ne donnent qu'un « aperçu » des conditions actuelles. On mène rarement des enquêtes longitudinales, en partie parce qu'elles sont extrêmement coûteuses. En permettant aux chercheurs d'utiliser les divers ensembles de données recueillis sur des sujets similaires, une fonction nationale d'archivage des données rendrait la recherche diachronique possible sans entraîner les coûts et les délais associés à de nombreuses années de collecte de données. En outre, le Canada aurait un outil pour suivre l'évolution de ses problèmes. Un service national d'archivage des données comblerait une importante lacune dans les domaines de la recherche et de l'enseignement supérieur.

#### **4.7 Modèles de conception pour les secteurs public et privé**

L'investissement approprié dans la conservation des données créées par ou pour les chercheurs *n'assurera pas que* l'accessibilité de ressources inestimables de recherche, il servira *aussi* de modèle à d'autres domaines — des sciences sociales au domaine de l'assurance — qui investissent dans la recherche.

L'archivage des données en tant que tel est une activité qui demande des recherches approfondies. Il y a un grand nombre d'aspects de l'archivage des données que nous ne savons pas comment mettre pratique, le plus important est la conservation à long terme, autant en termes d'aspects techniques que d'aspects conceptuels, comme les protocoles d'acquisition et de conservation. Partout dans le monde, les organisations, autant publiques que privées, se démènent pour conserver et gérer l'énorme volume de données créées quotidiennement. En répondant à ces besoins, des archives nationales de résultats de recherche seraient les seules à pouvoir mettre à contribution les talents en matière de recherche des informaticiens, des ingénieurs, des documentalistes, des bibliothécaires, des archivistes, des spécialistes en gestion, des juristes et des éthiciens ainsi que l'entière infrastructure bien élaborée des universités canadiennes.

#### **4.8 Accessibilité des chercheurs à des ressources coûteuses**

L'accessibilité est un élément clé pour les chercheurs. À l'heure actuelle, les conditions sont loin d'être idéales. Un grand nombre d'ensembles de données spatiales, par exemple, nécessaires à la recherche ne sont disponibles que selon un système de recouvrement des frais. Les coûts élevés d'ensembles de données et d'image de télédétection de ce genre les rendent inaccessibles à la plupart des chercheurs et même des bibliothèques parce qu'elles aussi doivent souvent payer pour y accéder. Il y a des cas où des établissements et des chercheurs canadiens doivent obtenir des données sur le Canada d'organismes américains parce qu'ils n'ont pas les moyens d'acheter les données de la source canadienne originale. On procède actuellement à une étude qui laisse entendre que les coûts de commercialisation des données se rapprochent du rendement financier obtenu. Les chercheurs et l'industrie ont indiqué avoir des doutes sur l'utilité de telles politiques. On devrait permettre l'accès public à des données de ce type au moyen d'un service national d'archivage des données. Si on ne le fait pas, le Canada pourrait très bien se retrouver avec une infrastructure nationale de données spatiales sans charpente.

*Le principe d'égalité d'accès devrait être observé peu importe le modèle d'archives de données recommandé. Le milieu universitaire se bat depuis longtemps pour éliminer les coûts qui ont entraîné des inégalités surprenantes dans le monde de la recherche, et même si on a conclu un arrangement avec Statistique Canada au sujet d'un niveau d'accès égal, un grand nombre de nos chercheurs n'ont toujours pas accès à des fichiers en raison de frais excessifs*

Réponse d'une personne interrogée

## 5.0 Conclusion

Comme l'ont indiqué avec justesse des membres du groupe de travail, la richesse du Canada en matière de résultats de recherche numériques est menacée. Même si cette déclaration semble exagérée, c'est une préoccupation que partagent un grand nombre d'archivistes, de bibliothécaires et de chercheurs du monde entier. L'information sous forme numérique est extrêmement fragile, et on peut en recueillir une grande quantité. Aujourd'hui, nous ne faisons que commencer à comprendre comment conserver, gérer et rendre accessible ces informations de manière efficace. Même s'il n'y a pas de raccourci facile pour régler les questions de la désuétude du support de données, des droits d'auteur et de la confidentialité, de la création de normes nationales et internationales ainsi que des limites de la culture de recherche actuelle, ne pas en tenir compte s'avérera coûteux à long terme.

Malgré nos investissements massifs en technologie de l'information, notre infrastructure de l'information et du savoir présente des faiblesses importantes qui s'accumulent. La réalité, c'est que nous perdons de précieuses ressources en matière de savoir parce qu'il n'y a pas d'organisme national mandaté pour en assurer l'administration. En conséquence, nous gaspillons inutilement les fonds publics.

La valeur des résultats de recherche canadiens réside dans leur utilité scientifique actuelle et future, l'investissement public consacré à la création des résultats de recherche et leur importante contribution à notre société. Sans la collecte volontaire, la conservation systématique, l'organisation intellectuelle et la gestion de nos résultats de recherche, nous perdrons une partie de notre patrimoine et minerons notre future capacité de recherche.

## 6.0 Recommandation

L'évaluation des besoins terminée, les membres des groupes de travail et d'étude recommandent à l'unanimité de passer à la deuxième phase de la consultation sur les archives nationales de résultats de recherche, où les membres examineront la meilleure approche en matière de système d'archivage national des résultats de recherche et proposeront un modèle de gestion efficace des résultats de recherche.

# Annexes

## Annexe 1

### Enquête auprès des chercheurs subventionnés par le CRSH

On a interrogé un échantillon<sup>3</sup> stratifié de 20 p. 100 de tous les chercheurs ayant reçu une subvention de soutien à la recherche du CRSH entre 1998 et 2000. Chaque chercheur a reçu un questionnaire portant sur son expérience de la création et de l'archivage de résultats de recherche ou de l'utilisation de résultats de recherche produits par d'autres. On leur a également demandé ce qu'ils pensaient des questions fondamentales à la base des principes de l'archivage des données et d'évaluer l'importance de la mise sur pied des services nationaux de conservation des résultats de recherche.

On a reçu<sup>4</sup> 116 réponses, ce qui correspond à un taux de réponse de 26 p. 100. On s'est efforcé d'assurer un taux de réponse équilibré entre les anglophones et les francophones. Vingt-deux pour cent de tous les titres de projets subventionnés entre 1998 et 2000 sont en français. Le taux de retour initial de la version française du questionnaire n'est que de 10 p. 100. Après avoir fait un suivi postal auprès des chercheurs dont le titre du projet de recherche est en français, ce pourcentage est passé à 22 p. 100.

Pour avoir un aperçu des personnes qui ont retourné le questionnaire, 64 p. 100 des intervenants ont obtenu leur plus haut grade d'études postsecondaires entre 1980 et 2000. Cette répartition est peut-être plus représentative du plus haut niveau de rendement universitaire des récents titulaires de subventions du CRSH que de la population globale de chercheurs. Néanmoins, la distribution du plus haut niveau de rendement universitaire couvre une période de 40 ans. La répartition par décennie est de 34 p. 100 dans les années 90, 30 p. 100 dans les années 80, 28 p. 100 dans les années 70 et 9 p. 100 avant 1970 (voir Tableau 4).

### Ampleur de la création et de l'archivage des données

Les conclusions de cette enquête nous aident à répondre à la question relative au degré auquel les chercheurs subventionnés par le CRSH créent et archivent des données. Premièrement, le nombre de chercheurs produisant des données issues de projets financés par le CRSH est important. Deuxièmement, les données d'un très petit nombre de ces projets sont archivées. En outre, les chercheurs qui indiquent avoir archivé des données donnent des sources qui ne sont pas des services authentiques d'archivage.

Le questionnaire commençait par une série de questions visant à déterminer le nombre de projets ayant entraîné la création de données. Premièrement, on a présenté une définition de résultat de recherche et des exemples.<sup>5</sup> On a ensuite demandé aux intervenants s'ils avaient créé des fichiers de données ou des bases de données dans leur plus récent projet de recherche financé par le CRSH qui correspondent à la définition et aux exemples donnés. Cinquante-cinq pour cent des chercheurs (voir Tableau 1) ont indiqué en avoir créé dans le cadre de leur plus récent projet. De ce groupe, seulement sept pour cent ont déclaré avoir déposé ces données auprès d'archives. Sur les 93 n'ayant pas déposé de données, 18 p. 100 ont indiqué avoir

<sup>3</sup> Voir l'Annexe 1 pour une description de la technique d'échantillonnage utilisée.

<sup>4</sup> Nombre en date du 6 avril 2001.

<sup>5</sup> « Par résultat de recherche, on fait référence aux informations numériques structurées de façon méthodique en vue de produire des nouvelles connaissances. » Deux exemples de résultats de recherche provenant d'une grande variété de résultats de recherche numérisés sont des fichiers contenant des informations numériques codées tirées de questionnaires et des fichiers codés à l'aide d'étiquettes représentant une certaine forme de structure. »

l'intention de le faire dans l'avenir. On a ensuite demandé aux chercheurs s'ils avaient créé des données dans le cadre d'anciens projets de recherche. Cinquante-neuf pour cent ont répondu en avoir créé et 18 p. 100 de ceux-ci ont mentionné les avoir archivées. Vingt-deux pour cent des intervenants ont indiqué avoir l'intention d'archiver les données de leurs anciens projets.

Ces résultats donnent suffisamment de renseignements pour estimer dans quelle mesure les chercheurs subventionnés par le CRSH produisent et archivent des données.

En calculant un intervalle de confiance de 95 p. 100 pour chacune des questions clés sur les données mentionnées plus tôt (voir Tableau 2), on en dérive une estimation de l'intervalle à l'intérieur duquel le nombre de chercheurs produisent et archivent des données. On a calculé des estimations prudentes des intervalles de confiance à l'aide du nombre de questionnaires retournés plutôt que de la taille originale de l'échantillon. Par exemple, l'intervalle de confiance du pourcentage de chercheurs qui ont créé des données dans le cadre de leur projet actuel financé par le CRSH (Q1A) est estimé à plus ou moins 4,5 p. 100 si on tient compte des 115 chercheurs qui ont répondu. L'intervalle varie entre 50,5 p. 100 et 59,5 p. 100. Exprimé en termes de taux par millier de chercheurs subventionnés par le CRSH, on obtient entre 505 et 595 chercheurs sur 1 000 qui produisent des fichiers de données et/ou des bases de données dans le cadre de leur projet.

<b>Tableau 1</b>				
<b>Répartition des chercheurs qui créent, archivent ou ont l'intention d'archiver des données</b>				
Question de l'enquête	Pourcentage de Oui	Nb		
		Réponse	Valide	Manquant
Q1A. Avez-vous créé des données ou une base de données dans le cadre de votre projet actuel subventionné par le CRSH?	55 %	63	115	1
Q1B. Avez-vous déposé ces données? (si non, passez à la question Q1F)	7 %	4	61	2
Q1F. Avez-vous l'intention de déposer ces données?	18 %	10	55	2
Q2A. Avez-vous créé des données ou une base de données lors de projets de recherche antérieurs?	59 %	68	116	0
Q2B. Avez-vous déposé ces données?	18 %	12	67	1
Q2D. Avez-vous l'intention de déposer ces données?	22 %	15	67	0

Si on utilise la taille de l'échantillon original pour calculer l'intervalle de confiance au lieu du taux de réponse valable, l'intervalle de confiance descend à plus ou moins 2,3 p. 100 ou environ la moitié des estimations du tableau 2. Dans pareil cas, le nombre de chercheurs produisant des données se situerait entre 530 et 570 sur 1 000. Au lieu de s'attarder à l'étendue de l'intervalle, on peut déduire d'après ces conclusions que des données sont créées dans un grand nombre de projets de recherche subventionnés par le CRSH. Les estimations les plus prudentes laissent entendre qu'entre 50 et 60 p. 100 des chercheurs créent des données. Si on applique ces estimations au nombre total de chercheurs subventionnés par le CRSH entre 1998 et 2000, le nombre de chercheurs produisant des données se situe entre 1 140 et 1 360.

Tableau 2						
Estimations de l'intervalle de confiance de 95 % pour les questions clés sur les données						
	N	p	q	IC	IC bas	IC élevé
Q1A Oui – données créées	115	0,55	0,45	0,045	50,5	59,5
Q1B Oui – données déposées	61	0,07	0,93	0,016	5,4	8,6
Q1F Oui – dépôt prévu	55	0,18	0,82	0,039	14,1	21,9
Q2A Oui – données créées	116	0,59	0,41	0,044	54,6	63,4
Q2B Oui – données déposées	67	0,18	0,82	0,035	14,5	21,5
Q2D Oui – dépôt prévu	67	0,22	0,78	0,041	17,9	26,1

Le fait que beaucoup de chercheurs produisent des données dans le cadre de leur recherche est compensé par la découverte alarmante que pratiquement aucune des données issues de ces projets n'est ou ne sera archivée. Sur les 55 p. 100 qui indiquent avoir créé des données, seulement 7 p. 100 les ont archivées (voir Tableau 1). Par ailleurs, lorsqu'on leur a demandé s'ils avaient déposé leurs données auprès d'archives, on a découvert que dans les faits, ils n'avaient pas tous déposé leurs données. Un des chercheurs a déposé ses données auprès d'un service de bibliothèque universitaire et un autre a envoyé une copie de ses données à la Bibliothèque nationale. Un a affiché ses données sur le Web, ce qui en soit ne constitue pas un dépôt d'archives et un autre a indiqué que si une initiative de la FCI était financée, il donnerait l'accès électronique aux données. Si on prend les estimations de l'intervalle de confiance du Tableau 2, le nombre de chercheurs qui archivent leurs données se situe entre 50 et 80 sur 1 000. Fait troublant, près de la moitié de ceux-ci n'auraient pas déposé leurs données auprès d'un service d'archivage.

Comparativement à ceux qui indiquent avoir archivé des données (Q1C), à peine un peu plus du double des chercheurs ont l'intention d'archiver les données issues de leurs projets (Q1F). Tel qu'indiqué dans le Tableau 1, 18 p. 100 révèlent avoir l'intention de déposer leurs données auprès d'un service d'archivage. On a ensuite demandé à ces chercheurs auprès de quelles archives ils déposeraient leurs données (Q1G). Deux sur dix ont indiqué des services universitaires de données, deux ne savaient pas où déposer leurs données, un a par erreur mentionné le centre de résultats de recherche de Statistique Canada, un a donné le nom d'un service d'archivage inexistant, un a indiqué qu'il créerait des archives si un service national n'était pas mis sur pied, un a mentionné l'ICPSR aux États-Unis, un a nommé des archives universitaires et un autre, un site Web. Sur les dix personnes qui ont répondu, moins de la moitié ont précisé un service d'archivage existant.

On a ensuite demandé aux chercheurs s'ils avaient créé des données dans au moins un de leur projet précédent (Q2A). Cinquante-neuf pour cent ont répondu oui. L'intervalle de confiance de 95 p. 100 pour cette estimation se situe entre 55 et 63 p. 100 des chercheurs. Le pourcentage de chercheurs soutenant avoir déposé des données d'un projet précédent (18 p. 100 – Q2B) est plus élevé que le pourcentage de ceux qui révèlent avoir déposé les données de leur projet actuel (7 p. 100 – Q1B). Les intentions de déposer les données sont légèrement supérieures pour les données des projets précédents (22 p. 100 pour la question Q2D comparé à 18 p. 100 pour la question Q1F), même si les intervalles de confiance de ces deux questions se recoupent, laissant entendre que les différences ne sont peut-être pas significatives.

Sur les 12 personnes ayant déclaré avoir déposé des données issues de recherches précédentes<sup>6</sup>, deux ont déposé leurs données auprès des archives de données du Royaume-Uni à Essex University. Quatre ont indiqué avoir donné des copies de leurs données à des

<sup>6</sup> Ces réponses sont tirées de la question Q2C du questionnaire destiné aux chercheurs.

bibliothèques universitaires. Ainsi, la moitié de ces intervenants ont signalé des services d'archivage existants. Deux ont répondu que les universités étaient des dépositaires. Une personne a mentionné les archives d'une université, mais a indiqué que celles-ci ont posé problème à la suite de changements au système informatique de l'université. À sa connaissance, aucune disposition n'avait été prise pour lire les données dont le format de fichier n'était plus utilisé. Un autre intervenant a mentionné le site Web de son centre tandis qu'un autre a mentionné son laboratoire de recherche. Enfin, un autre a indiqué avoir déposé une copie papier de ses données auprès d'un service d'archivage. Les trois dernières façons de faire ne sont pas des pratiques judicieuses d'archivage des données.

L'estimation de l'intervalle de confiance du nombre de chercheurs qui ont archivé des données issues de recherches antérieures se situe entre 145 et 215 chercheurs sur 1 000. D'après les lieux de dépôt rapportés, l'estimation des chercheurs comprend probablement des chercheurs qui ont déposé leurs données auprès d'un service d'archivage. Il est également probable que certains chercheurs n'aient pas déposé leurs données auprès d'un service d'archivage existant. La bonne nouvelle à l'effet que certains chercheurs ont déposé leurs données est assombrie par l'écart important entre le nombre de chercheurs qui ont créé des données par le passé et le nombre de ceux qui ont déposé leurs données auprès d'un service d'archivage. En termes de pourcentage, entre 80 et 85 p. 100 des chercheurs qui ont produit des données par le passé ne les ont pas déposées.

Vingt-deux pour cent des intervenants ont indiqué avoir l'intention d'archiver les données de leurs anciens projets (voir Tableau 1 – Q2D). Quarante pour cent des membres de ce groupe font partie des chercheurs qui ont indiqué avoir déposé des données issues d'anciennes recherches, c'est-à-dire les personnes ayant répondu oui à la question Q2B. Par conséquent, un pourcentage important de ceux ayant l'intention d'archiver leurs données ont déjà par le passé déposé des données. On leur a ensuite demandé auprès de quelles archives ils avaient l'intention de déposer les données de leurs précédentes recherches (Q2E). Étant donné que ceux-ci avaient déjà par le passé déposé des données, certaines réponses étaient similaires à celles données à la question Q2C. Trois des quinze intervenants ont indiqué qu'ils déposeraient leurs données auprès d'une informathèque universitaire. Deux ont mentionné qu'ils se fieraient aux archives universitaires, mais un a ajouté « à moins qu'il y ait une solution de rechange ». Deux chercheurs ont mentionné des lieux étrangers. Un a cité le Murray Research Center for the Study of Lives et le Radcliffe College de l'université Harvard et un autre, l'ICPSR. Trois ont déclaré être indécis ou ne pas savoir où déposer leurs données. Deux ont indiqué compter sur des sites Web. Deux autres voulaient trouver des archives appropriées et un a indiqué que tant qu'il n'y aurait pas de service d'archivage canadien, « nous aurons nos propres archives car nous ne sommes pas prêts à archiver nos données auprès d'archives étrangères ». Enfin, un intervenant a cité la Bibliothèque nationale ou le Musée canadien des civilisations. Ces huit dernières réponses révèlent qu'un peu plus de la moitié des personnes ayant répondu à cette question cherchent un endroit pour archiver leurs données ou ne savent pas ce que l'archivage englobe.

## Que pensent les chercheurs d'un service national d'archivage de données?

La majorité (60 p. 100) des chercheurs reconnaissent l'importance de mettre en place des services nationaux de conservation des résultats de recherche. Un bon nombre de chercheurs semble ne pas connaître pas les enjeux qui sous-tendent la conservation des résultats de recherche. Ces questions ne sont tout simplement pas abordées dans les universités canadiennes, que ce soit pendant la formation des étudiants diplômés ou dans les discussions concernant l'éthique professionnelle. Par ailleurs, des preuves indiquent qu'un grand nombre de chercheurs ne savent pas ce que l'archivage des résultats de recherche englobe. L'absence d'un service national voué à la conservation des résultats de recherche explique cette situation. Des résultats indiquent que diverses expériences de recherche semblent modeler les attitudes des chercheurs à l'égard du partage et de l'archivage des données. Par exemple, les chercheurs qui ont analysé les données recueillies par d'autres chercheurs sont plus favorables à la conservation des données.

On a tracé le profil des chercheurs d'après leur création ou leur utilisation des données. Catégoriser les chercheurs de cette manière nous a permis de définir les différences d'attitudes des chercheurs à l'égard de l'archivage des données. Il y a des chercheurs qui produisent et utilisent activement des données et d'autres qui n'en utilisent pas du tout. Comme le montre la figure 1, 76 p. 100 des chercheurs ayant répondu utilisent ou ont utilisé des données dans leurs recherches. Vingt-cinq pour cent de toutes les personnes interrogées ont répondu oui à trois questions — Q1A, Q2A et Q3A — au sujet de la création actuelle et passée de données et de l'utilisation de données produites par d'autres chercheurs. Un autre 29 p. 100 produisent des données à l'heure actuelle, en ont créé par le passé ou utilisent les données d'autres chercheurs. Vingt-deux pour cent n'ont pas produit de données dans le cadre de leur projet actuel, mais en ont produit par le passé et/ou utilisé qui appartenaient à d'autres chercheurs. Enfin, 24 p. 100 de ceux qui ont répondu non aux trois questions.<sup>7</sup>

L'année où on compte le niveau de scolarité le plus élevé varie entre trois types d'utilisateurs de données (voir les pourcentages dans le Tableau 4). Pour le groupe de non-utilisateurs de données, près du tiers ont obtenu leur diplôme le plus élevé avant 1980. Un autre tiers a obtenu son diplôme dans les années 80 et le dernier tiers, dans les années 90. La décennie pendant laquelle les non-utilisateurs de données ont suivi leur formation ne varie pas. Pour les deux catégories intermédiaires d'utilisateurs de données, leur répartition est d'environ 40 p. 100 avant 1980, 20 p. 100 dans les années 80 et 40 p. 100 dans les années 90. Les deux groupes intermédiaires ont reçu leur formation avant ou après les années 80. Enfin, la répartition du groupe d'utilisateurs de données le plus actif est de 31 p. 100 avant 1980 (personne n'a obtenu son diplôme le plus élevé avant 1970), 48 p. 100 dans les années 80 et 21 p. 100 dans les années 90. Près de la moitié de ce groupe a reçu sa formation dans les années 80.

On remarque des différences de cohortes intéressantes dans les décennies où les intervenants ont atteint le niveau de scolarité le plus élevé. Quarante pour cent de ceux de la cohorte qui ont obtenu leur diplôme dans les années 80 ont répondu oui aux trois questions relatives à l'utilisation des données (voir les pourcentages dans le Tableau 4). La cohorte des années 70 compte le pourcentage le plus important de chercheurs qui utilisent ou ont utilisé des données avec 84 p. 100. Le faible pourcentage de ceux de la cohorte des années 90 qui ont répondu oui aux trois questions sur l'utilisation des données (15 p. 100) peut s'expliquer par le fait que ces

---

<sup>7</sup> En utilisant ces quatre catégories, une répartition par type d'utilisateur de donnée et discipline d'étude est malheureusement impossible. On n'a pas demandé aux intervenants quel était leur champ d'étude. Ainsi, le pourcentage de non-utilisateurs de données en sciences humaines par rapport à ceux des sciences sociales ne peut être déterminé. Quelques-uns ont inscrit sur leur questionnaire que certaines questions les concernaient moins étant donné qu'ils sont en sciences humaines et non en sciences sociales. On ne peut cependant pas en conclure que tous les non-utilisateurs de données sont en sciences humaines.

chercheurs ne sont pas encore connus et par conséquent, n'ont pas eu la possibilité de produire des données par la passé. Enfin, ceux qui ont atteint le niveau de scolarité le plus élevé avant 1970 comptent le pourcentage le plus élevé de non-utilisateurs de données, même si ce groupe ne compte que 10 ayant répondu.

<b>Tableau 3</b>				
<b>Utilisation de données recueillies par d'autres (Q3A) et partage de données avec les autres (Q4)</b>				
Question de l'enquête	Pourcentage de Oui	Nb		
		Réponse	Valide	Manquant
Q3A. Avez-vous utilisé ou tenter d'utiliser des données produites par d'autres chercheurs?	39 %	45	116	0
Q3B. Vous a-t-on refusé l'accès à des données créées par d'autres chercheurs?	7 %	3	42	0
Q4. Avez-vous interdit à d'autres chercheurs l'accès à vos données?	8 %	9	114	2

En résumé, le modèle d'utilisation de données par les chercheurs qui ont atteint le niveau de scolarité le plus élevé au cours des trois dernières décennies indique qu'une proportion importante de ceux-ci produit et utilise des données. Quatre-vingt-quatre pour cent de la cohorte des années 70 fait partie de ce groupe. Soixante-quatorze pour cent de la cohorte des années 80 et 90 sont des utilisateurs actifs de données. Malheureusement, cette enquête ne permet pas de comparer l'utilisation des données entre les chercheurs des sciences humaines et ceux des sciences sociales car on ne leur a pas posé de question sur leur champ d'étude.

L'analyse secondaire a augmenté au cours des dernières années étant donné que de plus en plus de données deviennent accessibles grâce aux archives internationales de données, comme l'ICPSR, ou grâce aux services gouvernementaux d'abonnement aux données, comme l'Initiative de démocratisation des données. L'analyse secondaire fait prendre conscience aux chercheurs de la valeur des données recueillies par d'autres chercheurs. Pour évaluer dans quelle mesure les chercheurs ont recours à l'analyse secondaire, on leur a demandé s'ils ont déjà analysé des données recueillies par d'autres chercheurs. Trente-neuf pour cent de ceux qui ont répondu ont indiqué avoir utilisé ou tenté d'utiliser des données produites par d'autres chercheurs (Q3A). De ce groupe, sept pour cent ont déclaré n'avoir pu accéder aux données produites par d'autres (Q3B). Huit pour cent de tous ceux qui ont répondu ont indiqué avoir interdit à des chercheurs l'accès aux données qu'ils avaient produit (Q4). L'utilisation des données préparées par d'autres chercheurs semble être une pratique importante lorsqu'on examine les attitudes des chercheurs à l'égard du partage et de l'archivage des données. On donnera plus de détails plus loin.

On a utilisé 11 éléments (Q5A à Q5K) dans l'enquête pour évaluer l'attitude des chercheurs à l'égard des principes sous-jacents de l'archivage des données. Ces éléments concernent la légitimité de l'analyse secondaire comme méthode de recherche, la valeur des données en tant que sous-produit de la recherche, les questions de la propriété et du partage des données, le financement par le conseil de recherche de la préparation des données aux fins de partage et les répercussions des conseils de révision déontologique sur le partage des données (voir Liste 1). Étant donné que la formulation des cinq éléments (5a, 5d, 5e, 5h, 5j) n'appuie pas les principes

du partage ou de l'archivage des données, les catégories de réponse de ces éléments ont été exprimées pour correspondre avec l'orientation qui appuie ces principes.<sup>8</sup>

La figure 2 montre le pourcentage des intervenants qui sont « d'accord » ou « tout à fait d'accord » avec chaque élément. Les éléments de cette figure sont disposés en ordre décroissant d'appui. Quatre-vingt-un pour cent de ceux qui ont répondu sont d'accord avec l'énoncé selon lequel les données devraient être un sous-produit important de la recherche, tandis que seulement 21 p. 100 sont d'accord avec l'énoncé selon lequel les données n'appartiennent PAS au chercheur. L'ordre décroissant de l'accord représente une difficulté croissante à l'appui du partage et de la conservation des résultats de recherche. Six étapes des éléments de difficulté composent la figure 2. Les deux premiers éléments, c'est-à-dire les données en tant que sous-produit important et l'analyse secondaire en tant que méthode de recherche valable, reçoivent respectivement le soutien de 81 et de 78 p. 100 des intervenants. La deuxième étape des éléments portant sur la couverture par les conseils de recherche des coûts de préparation des données en vue de leur partage et sur le rôle des chercheurs comme administrateurs de données qui ne peuvent être facilement reproduites reçoit respectivement l'appui de 71 et de 68 p. 100 des intervenants. La troisième étape portant sur l'énoncé selon lequel les données devraient être partagées si on conserve l'anonymat du chercheur et sur l'énoncé selon lequel on doit sensibiliser les conseils de révision déontologique à la nécessité de conserver les données reçoit respectivement l'appui de 64 et 62 p. 100 des intervenants. Les deux éléments suivants constituent une étape légèrement plus importante. L'énoncé selon lequel affecter des ressources à la préparation des données issues de recherche ne constituerait pas du gaspillage reçoit l'appui de 50 p. 100 des intervenants et l'énoncé selon lequel l'archivage des données devrait faire partie intégrante du processus de recherche reçoit l'appui de 48 p. 100 de ceux-ci. À la cinquième étape, l'énoncé selon lequel les conseils de révision déontologique ne permettent pas le partage de données confidentielles sur les êtres humains reçoit l'appui de 28 p. 100 des intervenants, tandis que l'énoncé selon lequel les données ne devraient être partagées que si le chercheur principal y consent reçoit l'appui de 27 p. 100 de ceux-ci. Comme nous l'avons mentionné plus tôt, seulement 21 p. 100 sont d'accord avec l'énoncé selon lequel les données n'appartiennent PAS au chercheur principal.

On a fait un diagramme du nombre total d'éléments recueillant l'appui des intervenants (voir figure 4). Un pointage de zéro signifie que le chercheur n'appuie aucun des éléments favorables à l'archivage des données, tandis qu'un pointage de 11 signifie qu'il appuie tous les éléments. Comme le montre la figure 4, 17 p. 100 de ceux-ci sont peu favorables à l'archivage des données (ceux ayant un pointage entre zéro et trois), tandis que 24 p. 100 sont très favorables (ceux ayant un pointage entre 8 et 11). Cinquante-neuf pour cent se trouvent entre les deux. La corrélation entre ce diagramme et la question portant sur l'importance pour le Canada de mettre sur pied des services nationaux de conservation des résultats de recherche (Q6) est de 0,498. Cette corrélation relativement élevée corrobore l'interprétation selon laquelle ce diagramme mesure l'appui à l'archivage des données.

La figure 6 montre un tracé en boîte de ce diagramme pour chacun des quatre groupes d'utilisateurs décrits plus tôt. Manifestement, les chercheurs du groupe qui a répondu oui aux trois questions sur l'utilisation des données sont les plus favorables à l'archivage des données. La moyenne de ce groupe est de 7,5, tandis que l'autre moyenne la plus élevée (6,0) est celle du groupe de chercheurs qui n'ont pas créé de données dans le cadre de leur projet actuel, mais qui en ont créées dans le passé et/ou qui ont utilisé des données créées par d'autres chercheurs. Comme nous l'avons mentionné plus tôt, les chercheurs qui n'ont pas eu recours aux données créées par d'autres, mais qui en ont créé dans le cadre de leur projet actuel subventionné par le CRSH ainsi que par le passé semblent être le moins favorables à l'archivage des données. Par exemple, ce groupe a une moyenne de 5,0 sur l'échelle d'archivage des données. En outre, 89 p. 100 des membres de ce groupe sont d'accord ou tout à fait d'accord avec l'énoncé selon lequel les données appartiennent au chercheur principal.

<sup>8</sup> « Tout à fait d'accord » et « d'accord » sont les réponses qui appuient ces principes.

Un bon nombre de chercheurs semblent indécis en ce qui concerne les éléments de l'échelle d'archivage des données. La figure 3 donne le pourcentage des indécis. Plus de 25 p. 100 des intervenants sont indécis en ce qui concerne six éléments (5c, 5f, 5g, 5h, 5j, 5k), et deux éléments recueillent un peu moins de 25 p. 100 (5d et 5i). Les deux éléments au sujet des conseils de déontologie recueillent le pourcentage le plus élevé de réponses hésitantes avec 37 et 35 p. 100. Ces chiffres indiquent que beaucoup de chercheurs n'ont pas réfléchi aux questions représentées par ces éléments ou qu'ils sont mal renseignés.

Enfin, la figure 5 montre la répartition des chercheurs qui ont répondu à la question relative à l'importance pour le Canada de mettre sur pied des services nationaux de conservation des résultats de recherche (Q6). Soixante pour cent estiment que des services de ce genre sont importants ou très importants pour le Canada, tandis que seulement 12 p. 100 estiment qu'ils ne sont pas importants ou pas importants du tout. Vingt-huit pour cent sont incertains, ce qui correspond au pourcentage d'indécis à l'égard de plusieurs des éléments examinés plus tôt.

Le soutien à l'endroit de services nationaux varie grandement entre le type d'utilisateurs de données. Quatre-vingt-dix pour cent de ceux qui ont répondu oui aux trois questions relatives à l'utilisation des données conviennent de l'importance des services nationaux pour conserver les résultats de recherche. Soixante-douze pour cent de ceux qui n'ont pas créé de données dans le cadre de leur projet de recherche actuel, mais qui en ont créées dans le passé et/ou utilisées qui appartenaient à d'autres chercheurs sont du même avis. Le pourcentage d'appui des deux groupes restant chute considérablement, 44 p. 100 des non-utilisateurs de données conviennent de l'importance des services nationaux pour conserver les résultats de recherche et seulement 38 p. 100 des chercheurs qui produisent actuellement des données, en ont produit par le passé ou utilisé qui appartenaient à d'autres chercheurs sont du même avis. Ces deux derniers groupes comptent le pourcentage le plus élevé d'indécis avec 44 p. 100 chacun.

Le groupe d'utilisateurs de données comprenant les chercheurs qui produisent actuellement des données et qui en ont soit produit dans le passé ou utilisé qui appartenaient à d'autres chercheurs appuie le plus ardemment l'énoncé selon lequel les données appartiennent au chercheur principal (81 p. 100). Soixante-cinq pour cent des membres de ce groupe sont également d'accord avec l'énoncé à l'effet que les données ne doivent être partagées que si le chercheur principal y consent. Ces chercheurs semblent être très attachés aux données recueillies dans le cadre de leur projet de recherche subventionné par le CRSH.

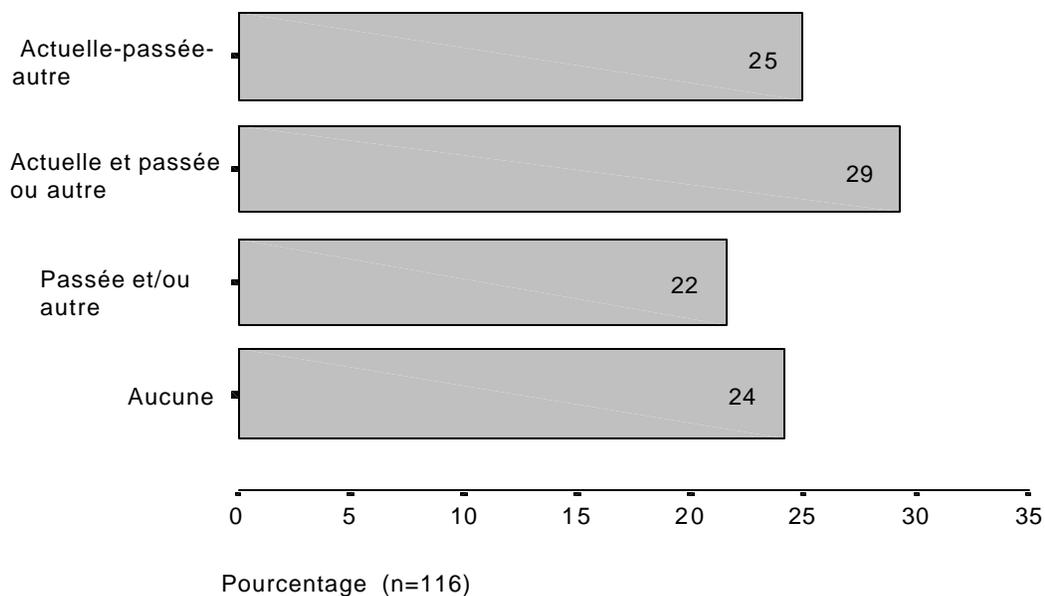
On remarque des différences entre l'accord avec la question Q6 et les cohortes de niveaux pendant la décennie où les intervenants ont atteint le niveau de scolarité le plus élevé. La cohorte qui considère le moins important des services nationaux de conservation de données est celle du groupe dont les chercheurs ont obtenu leur diplôme avant 1970, même si ce groupe ne compte que 10 personnes. Trente pour cent de cette cohorte appuie les services nationaux de conservation des données, tandis que 50 p. 100 sont incertains. L'autre cohorte la moins favorable à des services nationaux est celle composée de chercheurs ayant obtenu leur diplôme dans les années 90 avec 56 p. 100. Vingt-huit pour cent des membres de cette cohorte sont indécis. De la même façon, 29 p. 100 des membres de la cohorte ayant obtenu leur diplôme dans les années 80 sont incertains, mais 60 p. 100 considèrent importante la mise sur pied de services nationaux. La cohorte dont les membres ont obtenu leur diplôme dans les années 70 est celle qui considère le plus important la mise sur pied de services nationaux avec 74 p. 100. Seulement 19 p. 100 des chercheurs de cette cohorte sont indécis. Un tiers des membres de la cohorte ayant obtenu leur diplôme dans les années 90 est composé d'utilisateurs de données pour qui la mise sur pied de services nationaux de conservation des données est la moins importante (Q6), ce qui explique probablement le faible niveau d'appui de cette cohorte. Il faut quand même noter que 56 p. 100 de cette cohorte est favorable à la mise sur pied de services nationaux.

			Année d'obtention du diplôme d'études par type d'utilisateurs de données				Total
			Non- utilisateurs de données	Ont produit des données lors de précédentes recherches ou ont utilisé les données d'autres chercheurs	Producteurs actuels de données et qui ont soit produit des données auparavant ou ont utilisé les données d'autres chercheurs	Producteurs actuels de données, anciens producteurs de données et qui ont utilisé les données d'autres chercheurs	
Année	Avant 1970	Nbre	4	1	5		10
		Rangée %	40,0 %	10,0 %	50,0 %		100,0 %
		Col %	14,3 %	4,0 %	14,7 %		8,6 %
1970- 1979	1970- 1979	Nbre	5	9	9	9	32
		Rangée %	15,6 %	28,1 %	28,1 %	28,1 %	100,0 %
		Col %	17,9 %	36,0 %	26,5 %	31,0 %	27,6 %
1980- 1989	1980- 1989	Nbre	9	5	7	14	35
		Rangée %	25,7 %	14,3 %	20,0 %	40,0 %	100,0 %
		Col %	32,1 %	20,0 %	20,6 %	48,3 %	30,2 %
1990- 1999	1990- 1999	Nbre	10	10	13	6	39
		Rangée %	25,6 %	25,6 %	33,3 %	15,4 %	100,0 %
		Col %	35,7 %	40,0 %	38,2 %	20,7 %	33,6 %
Total		Nbre	28	25	34	29	116
		Rangée %	24,1 %	21,6 %	29,3 %	25,0 %	100,0 %
		Col %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %

Tableau 4

Figure 1

Expériences des chercheurs en matière de création de fichiers de données ou de bases de données ou d'utilisation de fichiers de données ou de bases de données créées par d'autres

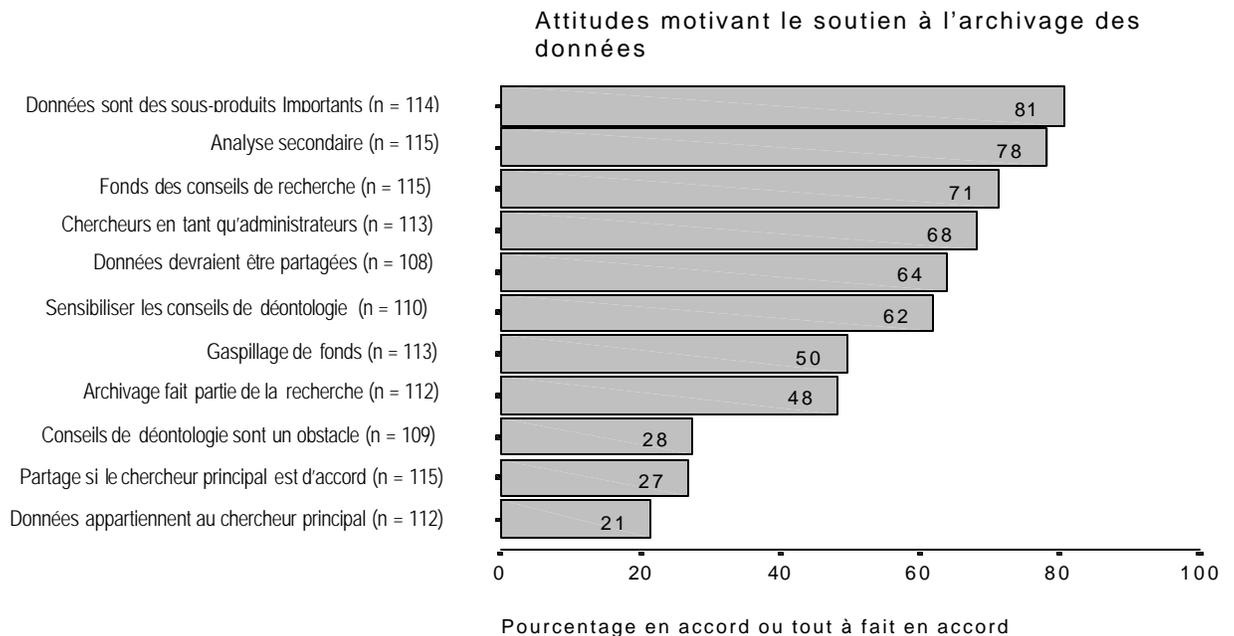


## Liste 1

### Éléments d'attitude expliquant l'appui à l'archivage des données

- 5a. L'analyse des données secondaires n'est pas une méthode de recherche valable.
- 5b. On devrait considérer les données comme un sous-produit important de la recherche.
- 5c. On devrait partager les données avec d'autres chercheurs si la confidentialité est respectée.
- 5d. Les données appartiennent au chercheur principal.
- 5e. Les données ne devraient être échangées que si le chercheur principal y consent.
- 5f. L'archivage des données devrait faire partie intégrante du processus de recherche.
- 5g. Les chercheurs qui obtiennent des informations ne pouvant pas être facilement reproduites sont, dans une certaine mesure, des administrateurs de données.
- 5h. Consacrer des ressources à la préparation des données issues de ma recherche afin que d'autres chercheurs puissent les utiliser serait du gaspillage.
- 5i. Les conseils de recherche devraient inclure des fonds pour couvrir les coûts de préparation des données aux fins du partage.
- 5j. Les conseils de révision déontologique ne permettent pas le partage de données confidentielles sur les êtres humains.
- 5k. On doit sensibiliser les conseils de révision déontologique à la nécessité de conserver les données.

Figure 2



## Attitudes expliquant l'appui à l'archivage des données

### Pourcentage d'indécis

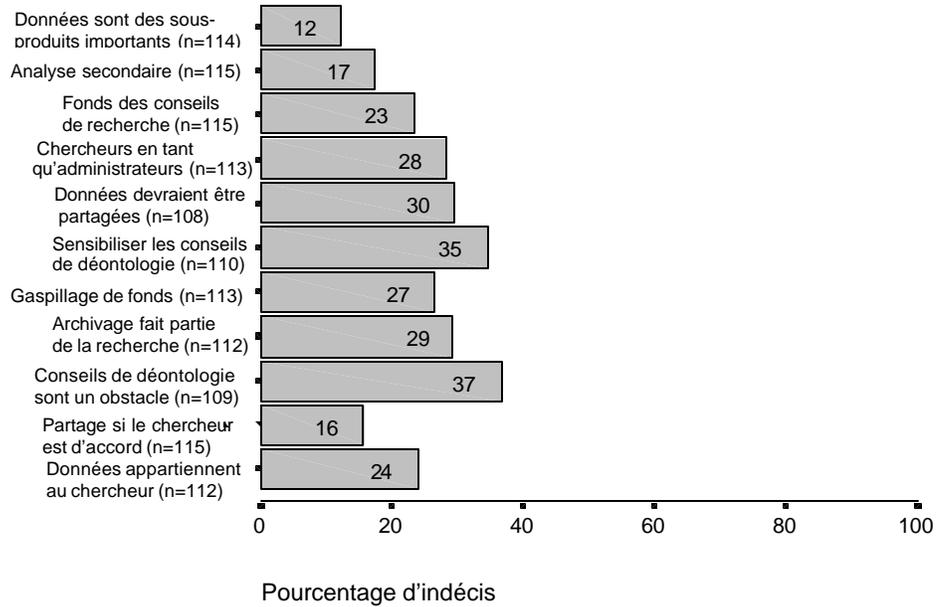


Figure 3

### Nbre d'éléments sous-jacents à l'archivage des données auxquels souscrivent les chercheurs

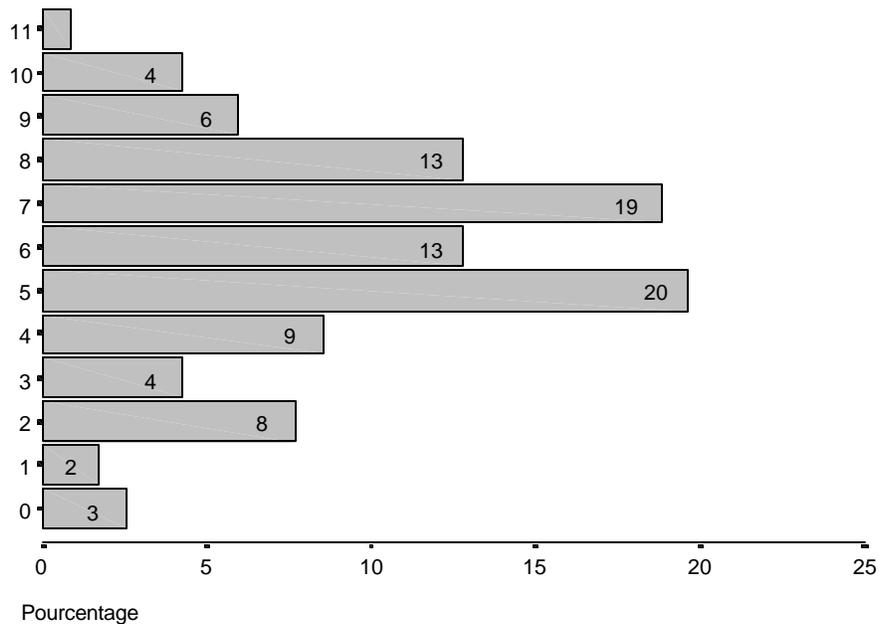


Figure 4

### Importance pour le Canada de créer des services nationaux de conservation des résultats de recherche

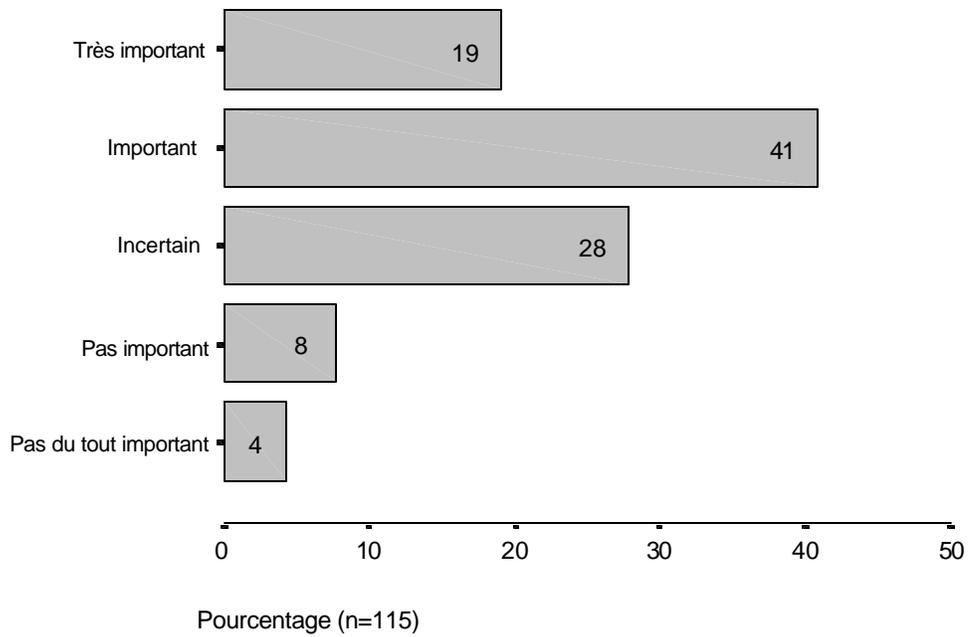


Figure 5

### Échelle d'attitude à l'égard de l'archivage par type de données

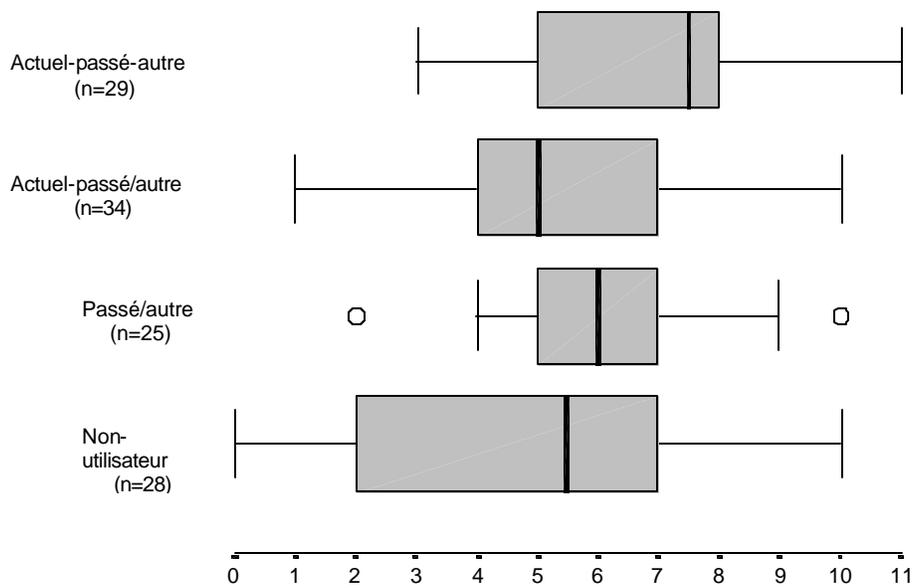


Figure 6

<b>Importance des services nationaux pour conserver les résultats de recherche (Q6) par type d'utilisateur de données et année d'obtention du diplôme d'études supérieures</b>									
Importance pour le Canada de créer des services nationaux pour conserver les résultats de recherche (Q6)*		Type d'utilisateur de données				Année d'obtention du diplôme d'études supérieures			
		Non-utilisateurs de données	Producteur de données par le passé et/ou utilisateur de données appartenant à d'autres	Producteur actuel de données et producteur de données par le passé ou utilisateur de données appartenant à d'autres	Producteur actuel de données, producteur de données par le passé et utilisateur de données appartenant à d'autres	Avant 1970	1970-1979	1980-1989	1990-1999
En désaccord	N	3	4	6	1	2	2	4	6
	Col %	11,1 %	16,0 %	17,6 %	3,4 %	20,0 %	6,5 %	11,4 %	15,4 %
Incertain	N	12	3	15	2	5	6	10	11
	Col %	44,4 %	12,0 %	44,1 %	6,9 %	50,0 %	19,4 %	28,6 %	28,2 %
En accord	N	12	18	13	26	3	23	21	22
	Col %	44,4 %	72,0 %	38,2 %	89,7 %	30,0 %	74,2 %	60,0 %	56,4 %

\* Les réponses « En total désaccord » ont été intégrées à « En désaccord » et « Tout à fait d'accord » à « D'accord ».

Tableau 5

Voici les réponses des chercheurs à une question ouverte au sujet des éléments examinés dans l'enquête :

- A. Question délicate — beaucoup d'aspects compliqués et de préoccupations éthiques doivent être examinés en profondeur avant de prendre une décision relativement au partage des données.
- B. En général, ce sont des questions auxquelles je n'ai pas été confronté. Les bases de données textuelles sur lesquelles je travaille ont été conçues pour mon usage personnel, et je ne crois pas qu'à l'état brut elles puissent être utiles aux autres, bien que j'espère que mes conclusions leur seront utiles. Si je travaillais en sciences sociales, j'aurais sans aucun doute un autre point de vue.
- C. On devrait encourager l'archivage des données sous forme numérique au lieu de le rendre obligatoire (p.ex. à l'aide de fonds supplémentaires réservés à cette fin et uniquement disponibles pour cet usage). On pourrait raisonnablement exiger des chercheurs qu'ils rendent leurs données accessibles à d'autres chercheurs sur demande (sous réserve d'exemptions raisonnables concernant la confidentialité), mais ce n'est pas comme les déposer dans des archives centrales.
- Je serais très inquiet si le CRSH imposait une technologie numérique uniforme (p. ex. certains programmes et/ou formats) comme exigence d'archivage, car en ce qui concerne l'informatique, on n'a qu'à penser à certaines exigences du CRSH en 1999 concernant les demandes de subvention en ligne.
  - Le pluralisme technologique devrait être inscrit à l'ordre du jour — Je suis content que vous ayez adopté cette approche avec ce questionnaire — nos systèmes de courriel semblent être incompatibles, je suis donc content d'avoir l'option télécopie/courrier ordinaire!
  - Je crois effectivement que les chercheurs ont l'obligation morale de partager des données recueillies en partie aux frais des contribuables une fois qu'elles sont publiées. Je me souviens avoir eu à faire à un chercheur réputé qui refusait même de divulguer son protocole de codage.
- D. Le Conseil pourrait commencer par un programme volontaire et accorder des fonds à des projets d'archivage des données à titre expérimental en faisant une révision après x temps.
- E. Je ne travaille pas dans un secteur où cette question pose problème à l'heure actuelle.
- F. Comme le laisse entendre la grande quantité de réponses « incertain », il y a des questions/problèmes auxquelles je n'ai pas été confronté, situation qui je le suppose est habituelle pour les chercheurs en sciences humaines.
- G. Ne s'applique à mon champ d'étude.
- H. Je ne fais pas ce genre de recherche.
- I. J'ai souvent pensé que les données (pas seulement les bases de données, mais aussi les notes de recherche sur disquette) devraient être conservées et que les chercheurs inactifs devraient être sensibilisés à la nécessité de remettre leurs disquettes à des archives universitaires ou publiques. J'estime que les données produites grâce au financement public devraient notamment être conservées.
- J. Comme vous pouvez peut-être le constater, je n'ai pas eu une bonne expérience avec des archives de données étrangères. Les procédures de ces archives ont changé entre-temps (du moins, en partie, à cause de ma mauvaise expérience et de celle d'autres chercheurs, en grande partie canadiens). À cet égard, elles peuvent revenir à leurs anciennes méthodes

(leur pratique actuelle est liée à leur source de financement, pas à leurs énoncés de politique), je ne suis pas prêt à courir le risque. Leur méthode consistait à facturer les chercheurs canadiens pour utiliser leurs propres données et celles d'autres chercheurs canadiens et à accorder une moins grande priorité à leur accès aux données qu'à celui des chercheurs de leur pays. En conséquence, je suis convaincu que le Canada doit avoir ses propres archives de données ou au moins être capable de conclure des accords nationaux avec les archives de données étrangères. Beaucoup de pays ont de tels accords avec les archives en question. En résumé, les chercheurs du pays où se trouvent ces archives ou ceux des pays ayant des ententes avec les archives en question n'avaient pas à payer pour accéder aux données et avaient la priorité d'accès à nos données canadiennes.

- K.
- On devrait rendre accessibles à tous les chercheurs les données issues de recherches subventionnées et pour lesquelles la confidentialité du chercheur est assurée.
  - La préparation de données aux fins d'utilisation par d'autres chercheurs améliore notre potentiel de création de connaissances et rend la création de données moins coûteuse (plus de chercheurs utilisent les données, moins les coûts sont élevés...).
  - Cette préparation est longue et coûteuse et doit par conséquent être subventionnée.
  - Des archives de données permanentes sont nécessaires pour stocker les données de façon sécuritaire, accroître les efforts de façon à renforcer l'utilisation secondaire des données et surtout promouvoir une culture de recherche préconisant le partage de données.
- L.
- On devrait exiger que les chercheurs subventionnés par les fonds publics archivent et mettent à la disposition des autres chercheurs les données produites, les instruments de recherche utilisés, les modèles utilisés et les cadres d'échantillonnage. On favoriserait ainsi la collaboration, les études longitudinales, les études de réplication, les études comparatives, la création de modèles de questions « normatives » dans certains domaines de recherche et les analyses secondaires. La transparence et la reddition de comptes seraient encouragées en exigeant l'archivage et l'accessibilité des données. En outre, la prise en compte de données de ce genre devrait devenir une caractéristique plus importante de la planification d'une « nouvelle » recherche primaire. Ainsi, on ne réinventerait pas la roue et il en résulterait peut-être des travaux plus créatifs. C'est du moins ce que j'en pense, pour ce que ça vaut.
- M.
- Comme je l'ai déjà mentionné, je crois qu'il est difficile de répondre à un grand nombre de ces questions par oui/non en accord/en désaccord puisque certaines questions sont plus complexes et que mes réponses varient en fonction de différents facteurs. Toutefois, malgré tout, je crois qu'il est effectivement important de conserver les résultats de recherche. J'estime qu'on doit se pencher sur la complexité des questions et des procédures.
- N.
- Étant donné que beaucoup de problèmes et de questions relatives à la structure, à l'éthique et à l'accessibilité des données concernent des disciplines ou des périodes précises, tenter d'élaborer une seule politique les englobant toutes me semble improductive. Les candidats à la recherche devraient peut-être se pencher sur les questions relatives à la conservation des données, et les évaluateurs et les comités, les intégrer à leurs évaluations sans imposer une politique uniforme et, au bout du compte, arbitraire. Selon moi, la structure et l'interprétation des données ne sont en aucun cas indépendantes l'une de l'autre; plus souvent qu'autrement, la distinction qu'on fait entre les données et l'interprétation est discutable.
- O.
- Les deux questions que je soulève ne sont pas directement liées à vos préoccupations à l'égard des documents numérisés.
1. Une grande partie de mon travail est consacrée à des entrevues avec des décideurs. Par le passé, je prenais des notes, mais je n'enregistrais rien. Si je reçois la subvention du CRSH demandée, je vais enregistrer et transcrire les entrevues effectuées dans le cadre

de mon projet. Dans ce cas-ci, il est plus logique de rendre ces entrevues accessibles à mes collègues, mais procéder ainsi est plus coûteux.

2. Je conserve mes données dans des boîtes à dossier. L'entreposage devient un problème, il faudrait que je loue de l'espace ou que la bibliothèque s'occupe d'entreposer mes documents. Ce problème est moins délicat que celui examiné par la présente étude, mais il concerne probablement plusieurs autres chercheurs et est pertinent pour la recherche historique dans l'avenir.

P. **NOTA** : J'ai eu du mal à répondre à certaines questions, comme celles des droits des chercheurs sur les données qu'ils produisent, par exemple. J'ai répondu selon l'opinion que j'ai de la situation en général, mais un grand nombre d'exceptions me viennent à l'esprit, comme celles relatives à la confidentialité, à la quantité de créativité et d'efforts nécessaires pour produire les données en question.

Q. La recherche à laquelle je participe à l'heure actuelle exigerait le consentement des sujets de recherche si je voulais archiver des données. Les données seront accessibles sous forme notamment de rapports qui feront partie d'un site Web. On dévoilera nos instruments de recherche. Les sujets de recherche recevront des copies de leur transcription d'entrevue. En bref, je crois qu'il y a différents moyens de partager des données selon le type de recherche entreprise.

R. Je me demande si cette question revêt plus d'importance dans certains domaines de recherche que dans d'autres. L'archivage des données ne semble pas être très important pour la recherche dans mon domaine, mais je peux comprendre l'importance qu'il peut avoir dans d'autres domaines. Néanmoins, je pense que le chercheur qui a recueilli les données doit avoir certains droits sur celles-ci.

S. À mon avis, les réponses à la question 5a. (l'analyse de données secondaires n'est pas une méthode de recherche fiable) et les questions suivantes doivent être nuancées. Je suis d'accord pour qu'un groupe de chercheurs puisse faire des analyses secondaires sur leur banque de données car ils/elles en connaissent les limites. Un groupe de chercheurs pourrait utiliser une banque de données d'un autre groupe de chercheurs seulement si la méthodologie de collecte des données était très simple, peu complexe.

Je ne crois pas que le conseil devrait consacrer beaucoup d'argent à la mise sur pied d'une structure pour forcer la création de banque de données publiques. L'analyse secondaire des données sera bonne seulement si les auteurs de la collecte de données participent à une seconde analyse. Le conseil pourrait appuyer financièrement la collaboration entre chercheurs pour une analyse secondaire de la banque de données.

T. Il me semble qu'il est extrêmement difficile de traiter comme un bloc homogène l'ensemble des données pouvant être dérivées de la recherche subventionnée. J'admets que certaines données sensibles devraient être maintenues confidentielles. Mais cela n'est certainement pas le cas de l'ensemble des données. En règle générale, des résultats scientifiques obtenus avec des fonds publics devraient appartenir au domaine public.

## Annexe 2

### Consultation nationale sur les archives de données Membres du groupe de travail et du groupe d'étude

#### Groupe de travail

**M. John ApSimon**

Adjoint spécial du président  
Carleton University

**M. José Igartua**

Département d'histoire  
*Université du Québec à Montréal*

**M. Gérard Boismenu**

Département de sciences politiques  
Université de Montréal

**M. Charles Humphrey**

Bibliothécaire  
University of Alberta

**M. Ian Lancashire**

Département d'anglais  
University of Toronto

**M<sup>me</sup> Sue Bryant**

**Secrétariat du Conseil du Trésor**

Coordinatrice principale de projet  
Secrétariat de l'infrastructure à clé publique

**M<sup>me</sup> Luciana Duranti**

École de bibliothécaires  
Études archivistiques et de l'information  
University of British Columbia

**M. Michael Murphy**

Directeur du centre Rogers de communications  
Ryerson University

**M. Matthew Mendelsohn**

Département des études politiques  
Queen's University

#### Groupe d'étude

**M. Paul Bernard**

Département de sociologie  
Université de Montréal

**M. Joseph Desloges**

Département de géographie  
University of Toronto

**M. Fraser Taylor**

Département de géographie  
Carleton University

**M. Geoffrey Rockwell**

Département des langues vivantes  
McMaster University

**M<sup>me</sup> Wendy Watkins**

Bibliothécaire  
Carleton University

**M. Michael Ridley**

Bibliothécaire en chef  
University of Guelph

**M. Ernie Boyko**

Directeur  
Bibliothèque et centre d'information  
Statistique Canada

**M. Frits Pannekoek**

Directeur  
Ressources de l'information  
University of Calgary

**M. Martin Brooks**

Chef d'équipe, Information interactive  
Institut de technologie de l'information  
Conseil national de recherches du Canada

**M<sup>me</sup> Yvette Hackett**

Agente des documents électroniques  
Division des archives gouvernementales et de  
la disposition des documents  
Archives nationales du Canada

**M. Douglas Hodges**

Services de technologie de l'information  
Bibliothèque nationale du Canada/  
Archives nationales du Canada

**M. Joseph R. Desloges**

Département de géographie et des  
programmes de planification  
*University of Toronto*

**M. Timothy Jackson**

Professeur agrégé de nouveaux médias  
Ryerson University

**M<sup>me</sup> Wanda M. Noel**

Avocate

## Annexe 3

### Consultation nationale sur les résultats de recherche Liste des présentations

#### Lettres des sous-ministres

**Peter Harrison**  
Sous-ministre  
Ressources naturelles Canada

**Shirley Serafini**  
Sous-ministre  
Affaires indiennes et du Nord Canadien

**David Dodge**  
Sous-ministre  
Santé Canada

**Alan Nymark**  
Sous-ministre  
Environnement Canada

#### Présentations des intervenants

1. **Association canadienne des utilisateurs de données publiques**
2. **M. A.W. Taylor**, président, Canadian Centre for Activity and Aging  
Facultés des sciences de la santé, de médecine et de dentisterie  
University of Western Ontario
3. **M. John Wilson**, département de science politique  
University of Waterloo
4. **M. Tom Nesmith**, professeur agrégé, département d'histoire  
Faculté des arts, St. Paul's College  
University of Manitoba
5. **M. Doug Hodges**, Services de technologie de l'information  
*Bibliothèque nationale du Canada*
6. **M. Michael Ornstein**, directeur, Institute for Social Research  
*Université York*
7. **D<sup>r</sup> Janice M. Morse**, directrice, International Institute for Qualitative Methodology  
Faculté des soins infirmiers, University of Alberta  
*Scientifique principale, Conseil de recherches médicales*
8. **M. Bryan Corbett**, président, Association of Canadian Archivists  
*et*  
**M. Fred Farrell**, président, Conseil canadien des archives
9. **M. Robert Garon**, conservateur  
*Archives nationales du Québec*

10. **M. Pierre Bordeleau**, vice-recteur adjoint aux TIC et directeur général  
et  
**M. Michel Lespérance**, secrétaire général  
**Direction générale des technologies de l'information et de la communication**  
**Université de Montréal**
11. **M<sup>me</sup> Miriam McTiernan**, Archiviste de l'Ontario  
Secrétariat du Conseil de gestion, Archives publiques de l'Ontario
12. **M<sup>me</sup> Elizabeth Krug**, agente de projet, Collections numérisées du Canada  
Industrie Canada
13. **M<sup>me</sup> Mary Jane Miller**, département des beaux-arts  
Brock University
14. **M. Marc Lacasse**, président  
*L'Association des archivistes du Québec*
15. **M. Douglas McLeod**, directeur de projet, NETERA Alliance  
Calgary, Alberta
16. **M. Gary Strike**, *bibliothécaire, services d'informathèque*  
University of Manitoba Libraries
17. **M<sup>me</sup> Rosemary Ommer**, directrice, Calgary Institute for the Humanities  
The University of Calgary  
et  
**M. Eric Sager**, président, département d'histoire  
University of Victoria

## Annexe 4

### Rapports sur l'archivage des résultats de recherche et autres documents connexes

- 1) Green, Ann, JoAnn Dionne et Martin Dennis. *Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata*, The Digital Library Federation, Council on Library and Information Resources, Washington, DC, juin 1999.
- 2) Hodge, Gail et Bonnie C. Carroll. *Digital Electronic Archiving: The State of the Art and the State of the Practice*, rapport financé par le Conseil international pour l'information scientifique et technique, comité de la politique de l'information, et CENDI, Oak Ridge, TN, avril 1999.
- 3) Haynes, David, David Streatfield, Tanya Jowett et Monica Blake. *Responsibility for Digital Archiving and Long Term Access to Digital Data*, David, Joint Information Systems Committee of the Higher Education Funding Council, Digital Archiving Working Group, Grande-Bretagne, 1997.
- 4) Société royale du Canada. *Data Policy and Barriers to Data Access in Canada: Issues for Global Change Research*, document de travail du Panel des systèmes de données et d'information du Programme canadien des changements à l'échelle du globe, 1996.
- 5) National Research Council. *Preserving Scientific Data On Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*, Steering Committee for the Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government, Commission on Physical Sciences, Mathematics and Applications, Washington, DC, 1995.
- 6) Gatenby, Pam. *Digital Archiving – Developing Policy and Best Practice Guidelines at the National Library of Australia*, National Library of Australia, janvier 2000.
- 7) Conseil international pour l'information scientifique et technique. *Digital Archiving: Bringing Issues and Stakeholders Together*, travaux de la conférence, janvier 2000, maison de l'UNESCO, Paris.
- 8) Hackett, Yvette. *An Overview of the Acquisition Policy of the National Archives of Canada*, Direction des documents gouvernementaux, Archives nationales du Canada, novembre 2000.
- 9) English, John. *Rôle des Archives nationales du Canada et de la Bibliothèque nationale du Canada*, rapport présenté à l'honorable Sheila Copps, Patrimoine canadien, gouvernement du Canada, 1998.
- 10) Hodges, Doug. *The National Library of Canada's Role in the Digital Environment*, Services de technologie de l'information, Bibliothèque nationale du Canada, novembre 2000.
- 11) Ryssevik, Jostein (service norvégien de données en sciences sociales) et Simon Musgrave (service d'archivage du Royaume-Uni). *The Social Sciences Dream Machine: Resource Recovery, Analysis and Delivery on the Web*, document présenté à la conférence de l'Association internationale pour les services et techniques d'information en sciences sociales, Toronto, mai 1999.
- 12) Chen, Su-Shing. « The Paradox of Digital Preservation », *Perspectives*, University of Missouri-Columbia, Institute of Electronic and Electrical Engineers, 2001.
- 13) Rieger, Y. Oya. « Digital Archiving: Approaches for Statistical Files, Moving Images, and Audio Recordings », *RGL DigiNews*, (Cornell University), vol.2, n° 6. décembre 1998.
- 14) Peterson, M.J. « Community and Individual Stakes in the Collection, Analysis and Availability of Data », *PS: Political Science and Politics*, septembre 1995, p.462-4.
- 15) King, Gary. « Replication, Replication », *PS: Political Science and Politics*, septembre 1995, p.444-52.
- 16) Economic and Social Research Council (Grande-Bretagne). *ESRC Green Paper on Data Policy and Data Archiving: Consultation Paper*, octobre 2000.

- 17) Lesk, Mihael. *Preserving Digital Objects: Recurrent Needs and Challenges*, Bellcore Corp., 2000.
- 18) Rothenberg, Jeff. *Ensuring the Longevity of Digital Information*, Santa Monica (Californie), Rand Corp., février 1999.
- 19) Buneman, Peter. *The Data that Archiving Fails to Capture*, University of Pennsylvania, 2000, [www.cis.upenn.edu/peter/archive.htm](http://www.cis.upenn.edu/peter/archive.htm)

CONSEIL DE RECHERCHES EN SCIENCES HUMAINES DU CANADA

## Savoir c'est comprendre

Le Conseil de recherches en sciences humaines du Canada (CRSH) est un organisme fédéral autonome qui appuie la recherche et la formation avancée en milieu universitaire dans les sciences humaines et en fait la promotion. Constitué par une loi du Parlement en 1977, le CRSH est dirigé par un Conseil de 22 membres qui rend compte au Parlement par l'entremise du ministre de l'Industrie.

La recherche subventionnée par le CRSH alimente des pensées novatrices portant sur des questions concrètes, notamment : l'économie, l'éducation, les soins de santé, l'environnement, l'immigration, la mondialisation, les langues, l'éthique, la paix, la sécurité, les droits de la personne, le droit, la pauvreté, les communications de masse, la politique, la littérature, la toxicomanie, la culture populaire, la sexualité, la religion, les droits des peuples autochtones, le passé et notre avenir.



Conseil de recherches en  
sciences humaines du Canada

Social Sciences and Humanities  
Research Council of Canada

Canada

350, rue Albert  
C.P. 1610  
Ottawa, ON K1P 6G4  
Canada

Tél. : 613.992.0691  
Télec. : 613.992.1787  
Internet : [www.sshrc.ca](http://www.sshrc.ca)