



DATA, DATA, EVERYWHERE... :

**Improving access
to population health
and health services
research data in Canada**

Final Report

April 2005

Submitted by:

**Charlyn Black and
Kimberlyn McGrail,
CHSPR, UBC**

**Cathy Fooks, Patricia Baranek and
Lisa Maslove,
Health Network, CPRN**



Centre for
HEALTH SERVICES AND POLICY RESEARCH

Canadian
Policy Research
Networks



Réseaux canadiens
de recherche en
politiques publiques

ISBN 1-897085-01-X

“Data Data Everywhere: Improving access to population health and health services research data in Canada”
was produced by:

Centre for Health Services and Policy Research
The University of British Columbia
426-2194 Health Sciences Mall
Vancouver, BC, Canada V6T 1Z3
Tel: (604) 822-1949
Fax: (604) 822-5690
Email: enquire@chspr.ubc.ca

Canadian Policy Research Networks Inc.
600-250 Albert St
Ottawa, ON K1P 6M1
Telephone: (613) 567-7500
Fax: (613) 567-7640
Email: publications_events@cprn.org

You can download this report from our websites at www.chspr.ubc.ca or www.cprn.org.

This publication is protected by copyright. It may be distributed for educational and non-commercial use, provided the Centre for Health Services and Policy Research and the Canadian Policy Research Networks are credited.

Table of Contents

THE CANADIAN POLICY RESEARCH NETWORKS.....	4
RÉSEAUX CANADIENS DE RECHERCHE EN POLITIQUES PUBLIQUES (RCRPP).....	5
THE CENTRE FOR HEALTH SERVICES AND POLICY RESEARCH.....	6
LE CENTRE FOR HEALTH SERVICES AND POLICY RESEARCH (CHSPR).....	7
ACKNOWLEDGEMENTS	8
REMERCIEMENTS	10
EXECUTIVE SUMMARY	12
Background.....	12
Views from data users and data collectors/custodians.....	13
Access.....	13
Electronic inventory	14
Recommended investments.....	14
Privacy and access issues	14
Existing inventory and data access activities.....	16
Building an inventory of databases.....	17
Next steps and recommendations.....	17
RÉSUMÉ	20
Contexte	20
Points de vue des utilisateurs, collecteurs et dépositaires de données	21
Accès	21
Inventaire électronique	22
Engagements recommandés	23
Enjeux associés à la protection de la vie privée et à l'accès	23
Inventaire existant et activités liées à l'accès aux données.....	26
Élaboration d'un inventaire de bases de données.....	26
Prochaines étapes et recommandations.....	27
CHAPTER 1: SETTING THE SCENE.....	30
1.1 The report.....	30
1.2 Background.....	31
References.....	34

CHAPTER 2: VIEWS FROM DATA USERS AND DATA COLLECTORS/CUSTODIANS.....	35
2.1 Introduction.....	35
2.2 Framework.....	35
2.3 Findings—Access.....	37
2.3.1 Availability.....	37
2.3.2 Accessibility.....	38
2.3.3 Accommodation.....	39
2.3.4 Affordability.....	41
2.3.5 Acceptability.....	41
2.3.6 Adequacy.....	42
2.4 Findings—Electronic inventory.....	44
2.5 Findings—Recommended investments.....	46
References.....	46
CHAPTER 3: PRIVACY AND ACCESS ISSUES IN THE USE OF POPULATION-BASED HEALTH AND HEALTH SERVICES DATA.....	47
3.1 Introduction.....	47
3.2 Organization of the literature review.....	48
3.3 Issues identified in the literature.....	48
3.3.1 Consent.....	48
3.3.2 Data linkage.....	50
3.3.3 Data retention and destruction.....	51
3.3.4 Security and safeguards.....	51
3.3.5 Review and oversight and the role of research ethics boards.....	51
3.3.6 Multiple rules, policies and procedures.....	52
3.3.7 Public communication.....	52
3.3.8 Legal and policy frameworks surrounding data access and privacy protection.....	53
3.4 Emerging developments and suggestions for future work.....	54
3.4.1 Develop privacy tool kits.....	55
3.4.2 Best Practice privacy guidelines or standards.....	55
3.4.3 Develop improved models of data stewardship.....	57
3.4.4 Strengthen and improve the practices of research ethics boards.....	57
3.4.5 Public communication about research and privacy.....	57
Bibliography.....	58
CHAPTER 4: EXISTING INVENTORY AND DATA ACCESS ACTIVITIES.....	62
4.1 Introduction.....	62
4.2 Methods.....	62
4.3 The review.....	62
4.3.1 Data and documentation in the United States.....	62
4.3.2 Data and documentation in the United Kingdom.....	64

4.3.3 Data and documentation in Canada.....	66
4.4 Overall findings	70
References.....	70
CHAPTER 5: BUILDING AN INVENTORY OF POPULATION HEALTH AND HEALTH SERVICES RESEARCH DATABASES FOR CANADA	71
5.1 Background	71
5.2 Methods.....	72
5.3 Tools developed.....	72
5.3.1 Conceptual framework and data collection tool.....	72
5.3.2 Tool to identify scope of databases for inclusion in an inventory	76
5.4 Findings, lessons learned and potential models.....	76
5.4.1 Lessons learned from developing and testing a prototype data collection tool.....	76
5.4.2 Potential models for future development of an inventory of databases.....	78
Bibliography.....	80
CHAPTER 6: NEXT STEPS AND RECOMMENDATIONS	82
6.1 Protecting privacy	82
6.2 Improving and increasing access to data	83
6.3 Developing an inventory of population databases	87
6.4 Concluding remarks	88
References	89
APPENDIX A: MISSION AND MANDATES OF CANADIAN AND INTERNATIONAL DATA AND RESEARCH ORGANIZATIONS	90
APPENDIX B: INTERVIEW METHOD	95
APPENDIX C: INTERVIEW INSTRUMENTS	97
APPENDIX D: SEARCH STRATEGY FOR ACCESS, PRIVACY AND CONFIDENTIALITY ISSUES	105
APPENDIX E: LITERATURE SEARCH STRATEGY AND OTHER METHODS FOR CHAPTERS 4 AND 5.....	111
APPENDIX F: DATA COLLECTION TOOL – FRAMEWORK FOR REVIEW OF DATABASES.....	116
APPENDIX G: PILOT TESTING THE DATA COLLECTION TOOL.....	124

The Canadian Policy Research Networks

Our mission is to create knowledge and lead public debate on social and economic issues important to the well-being of Canadians. Our goal is to help make Canada a more just, prosperous and caring society.

CPRN's trademark is its ability to help policymakers and citizens debate the beliefs, values, frameworks, policies, programs, and "ways of doing" that will help the country to cope with social and economic transformation.

CPRN fosters integration in a world that is increasingly fragmented by discipline, jurisdiction, language, and culture. It has unique process skills for shared learning, which shape the way research is performed and the way the results are communicated. It is a neutral space, where diverse groups of people can reflect, collaborate, and struggle with their differences in order to arrive at new understandings and to identify common ground.

CPRN's leaders are dedicated to generating constructive suggestions, based on strong analysis and a pragmatic understanding of what is possible in an imperfect world.

CPRN is independent. It is a non-profit organization with charitable status. It acquires its funding from diverse sources—federal and provincial governments, foundations, and corporations. This diversity ensures that no single voice dominates the research. The Board of Directors ensures good stewardship of these resources.

CPRN is cost-effective. Projects are ambitious in their scope, but costs and risks are spread across a number of funders. Overheads are minimized and start-up times are limited by attracting expertise from universities, think tanks, and other organizations. Dozens of people volunteer their time to participate in the governance and the research process.

Réseaux canadiens de recherche en politiques publiques (RCRPP)

La mission des RCRPP est de créer des connaissances et d'orienter le débat public sur des questions économiques et sociales qui ont une importance déterminante pour le bien-être des Canadiens. Leur objectif est de faire du Canada une société plus humaine, plus juste et plus prospère.

Les RCRPP se distinguent par leur capacité d'aider les décideurs et les citoyens à discuter de croyances, de valeurs, de cadres, de politiques, de programmes et de « façons de faire » qui aideront le pays à faire face aux transformations sociales et économiques.

Dans un monde de plus en plus fragmenté par discipline, pouvoir, langue et culture, les RCRPP encouragent l'intégration. Ils possèdent des compétences uniques favorisant la diffusion de connaissances, qui orientent la façon de conduire des recherches et de communiquer les résultats qui en découlent. Il s'agit d'un forum impartial par l'entremise duquel divers groupes de personnes peuvent réfléchir sur leurs différences et en venir à bout, collaborer pour acquérir une compréhension nouvelle de celles-ci et définir les points qu'ils ont en commun.

Les dirigeants des RCRPP font tout leur possible pour formuler des suggestions qui sont constructives en s'appuyant sur de solides analyses et une compréhension pragmatique de ce qui est possible de faire dans un monde imparfait.

Les RCRPP sont un organisme indépendant, à but non lucratif et à vocation de bienfaisance. Ils puisent leurs fonds de différentes sources, soit des gouvernements fédéral et provinciaux, des fondations et des sociétés. Puisque les bailleurs de fonds sont multiples, aucun d'entre eux ne peut exercer une domination dans la recherche. Le conseil d'administration veille à la saine gestion de ces ressources.

Les RCRPP sont rentables. Les projets sont ambitieux, mais ce sont les bailleurs de fonds qui en assument collectivement les coûts et les risques. En faisant appel à des spécialistes issus des universités, de groupes de réflexion et d'autres organisations, les frais généraux sont réduits au minimum et les périodes de démarrage sont courtes. Des douzaines de personnes donnent de leur temps pour participer aux processus de gestion et de recherche.

The Centre for Health Services and Policy Research

The Centre for Health Services and Policy Research (CHSPR) is an independent research centre based at the University of British Columbia. CHSPR's mission is to stimulate scientific enquiry into issues of health in population groups, and ways in which health services can best be organized, funded and delivered. Our researchers carry out a diverse program of applied health services and population health research under this agenda.

CHSPR aims to contribute to the improvement of population health by ensuring our research is relevant to contemporary health policy concerns and by working closely with decision makers to actively translate research findings into policy options. Our researchers are active participants in many policy-making forums and provide advice and assistance to both government and non-government organizations in BC, Canada and abroad.

CHSPR receives core funding from the BC Ministry of Health Services to support research with a direct role in informing policy decision-making and evaluating health care reform, and to enable the ongoing development of the BC Linked Health Database. Our researchers are also funded by competitive external grants from provincial, national, and international funding agencies.

Much of CHSPR's research is made possible through the BC Linked Health Database, a valuable resource of data relating to the encounters of BC residents with various health care and other systems in the province. These data are used in an anonymized form for applied health services and population health research deemed to be in the public interest.

CHSPR has developed strict policies and procedures to protect the confidentiality and security of these data holdings and fully complies with all legislative acts governing the protection and use of sensitive information. CHSPR has over 30 years of experience in handling data from the BC Ministry of Health and other professional bodies, and acts as the access point for researchers wishing to use these data for research in the public interest.

For more information about CHSPR, please visit www.chspr.ubc.ca.

Le Centre for Health Services and Policy Research (CHSPR)

Le CHSPR est un centre de recherche indépendant qui est basé à l'Université de la Colombie-Britannique. Sa mission est de stimuler les enquêtes scientifiques sur des enjeux touchant la santé chez des groupes de population et de trouver des façons d'améliorer l'organisation, le financement et la prestation des services de santé. Nos chercheurs dirigent un programme diversifié de recherche appliquée sur les services de santé et la santé des populations selon les objectifs établis par le CHSPR.

L'objectif du CHSPR est de contribuer à améliorer la santé des populations en veillant à ce que les recherches qui s'y font correspondent aux préoccupations actuelles en matière de politique de la santé et en travaillant étroitement avec les décideurs pour transformer activement les résultats de recherche en options stratégiques. Nos chercheurs participent activement à de nombreux forums directeurs et donnent des conseils et de l'aide aux organismes gouvernementaux et non gouvernementaux en Colombie-Britannique, ailleurs au Canada et à l'étranger.

Le CHSPR reçoit son financement de base du ministère des Services de santé de la Colombie-Britannique pour soutenir la recherche et a comme rôle direct d'informer ceux qui prennent des décisions stratégiques, d'évaluer la réforme des soins de santé et de permettre l'élaboration continue de la BC Linked Health Database (base de données liée sur la santé de la Colombie-Britannique). Nos chercheurs reçoivent également des fonds en participant à des concours de subvention externes, qui sont octroyés par des organismes de financement provinciaux, nationaux et internationaux.

Une grande partie de la recherche effectuée au CHSPR est rendue possible grâce à la BC Linked Health Database, une source précieuse de renseignements sur les expériences ayant été vécues par les Britanno-colombiens dans le système de soins de santé et dans d'autres systèmes de la province. Ces données sont présentées anonymement et servent à conduire des recherches appliquées sur les services de santé et la santé des populations que l'on juge d'intérêt public.

Le CHSPR a mis au point des politiques et des procédures strictes afin de respecter la confidentialité des fonds de données et d'en assurer leur protection ainsi que de suivre à la lettre toutes les lois régissant la protection et l'utilisation d'information sensible. Depuis 30 ans, le CHSPR traite les données du ministère des Services de la santé de la Colombie-Britannique et d'autres organismes professionnels. Il est le point d'accès des chercheurs souhaitant se servir de ces données pour faire des recherches visant l'intérêt du public.

Pour obtenir plus d'information sur le CHSPR, visitez son site Web à www.chspr.ubc.ca (en anglais).

Acknowledgements

This project was sponsored by the Canadian Institutes of Health Research (CIHR)—Institutes of Population and Public Health, Health Services and Policy Research, Aboriginal Peoples’ Health, Aging, Cancer Research, Circulatory and Respiratory Health, Gender and Health, Genetics, Infection and Immunity, Musculoskeletal Health and Arthritis, Neurosciences, Mental Health and Addiction, and Nutrition, Metabolism and Diabetes—with the Canadian Population Health Initiative (a part of the Canadian Institute for Health Information), Health Canada’s Centre for Surveillance Coordination and Statistics Canada. CIHR’s Institute of Population and Public Health, located at the University of Toronto, served as the administrative lead organization for this project.

The research team gratefully acknowledges the contributions of many people to the development of this report.

A Steering Committee, ably chaired by Diane Watson, provided guidance throughout the course of this project and pointed us to relevant initiatives and expertise. Thank you to: Erica Di Ruggiero, Daniel Friedman, Elizabeth Gyorfi-Dyke, Mary Catherine Lindberg, David Loukidelis, Vera Ndaba, Gregory Sherman and Michael Wolfson.

Michele O’Rourke gave us excellent secretariat support, organizing and minuting steering committee meetings. Kim Gaudreau ably took these tasks over toward the end of the project.

A series of key informant interviews provided us with in-depth understanding of many current and ongoing initiatives related to data development, access, custodianship and inventories. Interviewees included: Richard Alvarez, Nick Black, John Frank, Chuck Humphrey, Louise Ogilvie, Greg Sherman, Steve Slade, Diane Watson, Greg Webster, Michael Wolfson, Glenda Yeates and Jennifer Zelmer.

Over forty anonymous interview participants provided useful insights into the worlds of data collection, custodianship, access and use.

Lars Apland provided able research assistance with the development of the conceptual framework. Mary-Doug Wright and Catherine Howett identified and retrieved resource material. Peter Schaub produced the figures in Chapter 5 of the report.

Michele Wiens, Karen Hofmann and Carole Herbert provided expert assistance in the testing of our conceptual framework/data collection tool.

Three external reviewers—Carolyn DeCoster, Steven Lewis and Don Willison—offered thoughtful critiques of an early draft of the report that helped us refine and in some cases re-frame our findings and recommendations. We are also grateful to Judith Maxwell, president of CPRN, whose insightful comments helped us think about the governance issues surrounding our recommendations. Heidi Matkovich provided invaluable editing of an early draft and the final version of the report.

Cathy Fooks was the Principal Investigator on this project until June 30, 2004 in her previous role as Director of CPRN’s Health Network. As of September 2004, Tom McIntosh stepped into this role and provided helpful input and guidance as we finished the report.

The conclusions and opinions found in the report are solely those of the authors and therefore do not necessarily represent the views of the funding partners (CIHR, CIHI-CPHI, Statistics Canada and Health Canada) and the members of the steering committee. Thus no official endorsement by the funding partners or the members of the steering committee is intended or should be inferred.



Standards development for Public health



Remerciements

Ce projet a été financé par les Instituts de recherche en santé du Canada (IRSC), dont : l'Institut de la santé publique et des populations, l'Institut des services et des politiques de la santé, l'Institut de la santé des Autochtones, l'Institut du vieillissement, l'Institut du cancer, l'Institut de la santé circulatoire et respiratoire, l'Institut de la santé des femmes et des hommes, l'Institut de génétique, l'Institut des maladies infectieuses et immunitaires, l'Institut de l'appareil locomoteur et de l'arthrite, l'Institut des neurosciences, de la santé mentale et des toxicomanies et l'Institut de la nutrition, du métabolisme et du diabète, avec l'aide de l'Initiative sur la santé de la population canadienne (laquelle fait partie de l'Institut canadien d'information sur la santé [ICIS]), le Centre de coordination de la surveillance de Santé Canada et Statistique Canada. L'Institut de la santé publique et des populations des IRSC, qui est situé à l'Université de Toronto, a assumé la responsabilité administrative du présent projet.

L'équipe de recherche souhaite sincèrement remercier bon nombre de personnes qui ont contribué à la rédaction de ce rapport.

Le comité directeur, dirigé avec compétence par Diane Watson, ont fourni des avis tout au long de l'exécution du projet, et nous ont informées de compétences et d'initiatives connexes. Nous remercions les personnes suivantes : Erica Di Ruggiero, Daniel Friedman, Elizabeth Gyorfi-Dyke, Mary Catherine Lindberg, David Loukidelis, Vera Ndaba, Gregory Sherman et Michael Wolfson.

Michèle O'Rourke nous a fourni des services de secrétariat hors pair en organisant les réunions du comité directeur et en rédigeant les procès-verbaux. C'est Kim Gaudreau qui a habilement pris la relève vers la fin du projet.

Nous avons fait une série d'entrevues clés qui nous ont permis d'approfondir nos connaissances sur de nombreuses initiatives qui sont en cours traitant de l'élaboration de données, de leur accès et garde ainsi que d'inventaires. Parmi les personnes interviewées, notons les personnes suivantes : Richard Alvarez, Nick Black, John Frank, Chuck Humphrey, Louise Ogilvie, Greg Sherman, Steve Slade, Diane Watson, Greg Webster, Michael Wolfson, Glenda Yeates et Jennifer Zelmer.

Plus de quarante personnes ont été interviewées sous le couvert de l'anonymat. Celles-ci nous ont présenté des points de vue qui nous ont été utiles au sujet de la collecte de données, de leur garde, accès et utilisation.

Lars Aplan a agi avec compétence à titre d'adjoint à la recherche dans l'élaboration d'un cadre conceptuel. Mary-Doug Wright et Catherine Howett ont cherché du matériel documentaire et l'ont récupéré. Peter Schaub a produit les figures présentées dans le chapitre 5 du rapport.

Michele Wiens, Karen Hofmann et Carole Herbert ont aidé à tester le cadre conceptuel et l'outil de collecte de données.

Il y a eu trois examinateurs externes, soit Carolyn DeCoster, Steven Lewis et Don Willison, qui nous ont présenté de judicieux commentaires sur la première ébauche du rapport, nous permettant ainsi de retravailler nos conclusions et nos recommandations et, dans certains cas, de les reformuler. Nous sommes également reconnaissantes envers Judith Maxwell, la présidente des RCRPP, pour ses précieux commentaires, lesquels nous ont incitées à réfléchir aux questions de gérance liées aux recommandations. Heidi Matkovich nous a donné des services de rédaction inestimables puisqu'elle a produit la première ébauche du rapport et sa version définitive.

Cathy Fooks a été la chercheuse principale du projet jusqu'au 30 juin 2004 lorsqu'elle occupait son ancien rôle de directrice du Réseau de la santé du RCRPP. Depuis septembre 2004, c'est Tom McIntosh qui a pris la relève et qui nous a donné des suggestions et des conseils au moment de mettre la touche finale au rapport.

Les conclusions et les opinions formulées dans le présent rapport sont celles des auteurs. Elles ne reflètent pas nécessairement le point de vue des partenaires financiers (IRSC, Initiative sur la santé de la population canadienne de l'ICIS, Statistique Canada et Santé Canada) ni celui des membres du comité directeur. Il ne faudrait ainsi pas conclure ou prétendre que les partenaires financiers et les membres du comité directeur appuient ces conclusions et opinions.



Standards development for Public health



Executive Summary

In the fall of 2002, the Canadian Institutes of Health Research (CIHR) Institutes of Population and Public Health and Health Services and Policy Research jointly issued a request for proposals (RFP) with the Canadian Population Health Initiative (a part of the Canadian Institute for Health Information), Health Canada's Centre for Surveillance Coordination and Statistics Canada. The objectives of this RFP were to describe the current status of population-based health and health services databases in Canada and to show the potential for their use in innovative and important health research. For the purposes of this project, population-based health and health services data were defined as administrative databases, registries and survey databanks that are representative of an entire population who reside in a geographic region. The RFP noted that while Canada has some of the best-developed data repositories for studying health and health care, "the challenge now lies in enhancing access to and use of current data infrastructure for the purposes of conducting important health research and to make wise investments to increase data and analytic capacity."

The Canadian Policy Research Networks (CPRN), in partnership with the Centre for Health Services and Policy Research (CHSPR) at the University of British Columbia, was awarded the contract to undertake this project. This report presents the results of interviews with data collectors, custodians and users to identify current or emerging issues around collection, storage and use of data; reviews the current landscape of privacy and access issues in Canada; surveys international and Canadian activities in providing information about and access to data sets; outlines considerations for creating an inventory of population health and health services research databases; describes a prototype data collection tool that could assist in the development of such an inventory; and makes recommendations for moving forward the agenda of improving access to and use of Canadian data in the areas of population health and health services research.

The breadth of work undertaken meant that we were able to identify and offer recommendations for major topics, but we were not able to follow every concern in great detail. The broader context was also evolving as we conducted this study, at a speed that made it impossible for us to include everything that might be relevant or of interest. While we were not able to suggest immediate solutions in all cases, or to provide complete coverage of all topics, we hope this report provides useful direction for actions that can be taken to support researchers in population health and health services research in Canada.

Background

The current data environment in Canada involves a large number of players, all with differing mandates and roles. There are many data custodians in Canada, both national and provincial/territorial, that collect and maintain a wide range of population-based health and health services data.

However, these collectors and custodians, including Statistics Canada, the Canadian Institute for Health Information (CIHI), Health Canada, and provincial/territorial government departments and ministries, typically have no explicit mandate to support the research community. Data are collected for public health and surveillance, or, more commonly, in the course of operating health, education and social systems.

Regardless of the original purpose of collection, secondary analysis of such data has great potential for improving our understanding of the impact of public policy and other interventions on individuals and populations. But secondary analysis also necessitates a complex set of arrangements to govern the retention, disclosure and use of data. These issues become particularly contentious when research is not recognized as a primary mandate for the collecting agency.

As a consequence, Canada has a complex, fragmented set of arrangements by which some researchers can obtain access to data for research and others cannot. Interviews with data users and data collectors/custodians highlighted these and other issues.

Views from data users and data collectors/custodians

Forty-three interview respondents from across the country were asked to: identify the issues of greatest concern to them; consider whether they thought an electronic inventory of Canadian population-based health and health services databases would be a useful tool; and to nominate areas for future investments.

Access

Access quickly emerged as the major concern of data users, and responding to researcher requests for access as the major concern of data collectors/custodians. In organizing the observations and suggestions of interview respondents, we used a well established framework that describes several dimensions of access, each of which reflects the issues, challenges and opportunities facing data users and data collectors/custodians in the development, maintenance and use of population-based health and health services databases.

1. Availability - Research agencies should facilitate discussions between researchers/users and collectors/custodians to improve access to data. In addition, resources should be allocated to encouraging and creating greater linkages across databases and across jurisdictions.

2. Accessibility - Users suggested that every province should have a Statistics Canada Research Data Centre, and further recommended the creation of more provincial data centres within each province. Some custodians are currently exploring ways in which to extend data access through the liberalization of access policies, or the re-negotiation of licensing agreements.

3. Accommodation - Some users called for a pan-Canadian vision, a uniform, standardized policy for data access and a move away from multiple ad hoc arrangements. Others suggested the creation of a single data repository within each province whose sole purpose would be to ensure access to quality data from all government sectors. Other suggestions included the development of a universal format in which data are exported and better data documentation to make it easier for users to understand and use data. From the collector/custodian point of view, suggestions included liberalizing licensing agreements for centres in receipt of provincial government data. Most suggested that funding be increased to specifically account for the provision of data as one of their core businesses.

4. Affordability - Users recommended that public data should be free of charge and (subject to suitable privacy/confidentiality controls) available on the internet as it is in the US. Collectors/custodians recommended multi-year funding, and the recognition of the need to fund costs of data maintenance and cleaning. Technologies such as the electronic health record were seen as innovations that would decrease the work of data technicians in updating and cleaning records.

5. Acceptability - Users recommended an education program to address the public's concerns about privacy and security and to demonstrate the value of research in improving the overall health of the population and the functioning of the health care system. Some believe that it would assist public understanding if privacy legislation distinguished between *bona fide* researchers and other data users such as commercial organizations. Others recommended that the CIHR could facilitate discussions with collectors/custodians with the view to negotiating greater flexibility in access to data while providing assurances about protection of privacy and confidentiality. Overall, both users and collectors/custodians reported a need to better communicate and to build relationships and trust between each other.

6. Adequacy - Users and collector/custodians recommended that data quality should be recognized as a priority and should be reflected in the resources apportioned to it. Standardization of data definitions and collection methods, increased training of personnel, education of data users, the development of better data documentation, and the introduction of technologies such as telephone-assisted surveys and electronic health records would all greatly improve data quality.

Electronic inventory

In response to a question about the utility of an electronic inventory of population databases to support population health and health services research, the majority of both users and collectors/custodians provided either support or qualified support for development of such a resource. If an inventory were to be developed, many respondents suggested that it should be user-friendly, web-based with a dedicated website, and searchable by key words and standardized variables. Many felt it would be useful to have links to the data custodian's website, data dictionaries, data documentation, as well as links to articles/reports that used the data set. A number of respondents said an inventory should also provide a web-based portal to enhance actual data access.

Most believe that the custodian of the inventory should be an independent, national body, either a new body with this single mandate created through a federal/provincial/territorial agreement or an existing national body, such as CIHI, the CIHR Institutes of Health Services and Policy Research and Population and Public Health, or Statistics Canada.

Recommended investments

Respondents had many suggestions for the creation of new data sets, including:

- Health services data—community care, mental health, public health, drugs
- Population health data—chronic diseases,

disease staging data

- Biological and physical measurement data
- Longitudinal data—seniors, children, cohort studies
- Special populations—Aboriginal, homeless

Other suggested areas of investment included:

- negotiating standardized access and privacy policies;
- stable and ongoing funding for database maintenance and purchase;
- standardization of existing data sets and creation of data documentation;
- increased training for researchers to use large data sets and conduct secondary analysis, and for technicians to support and maintain large data sets;
- facilitating inter-regional comparisons and data linkage especially between health services data and determinants of health data;
- facilitating better understanding between collectors/custodians and users/researchers; and
- the development of a national vision and strategy for the collection, maintenance, and sharing of publicly funded Canadian data.

Privacy and access issues

The research environment is increasingly complex both because of rising public concerns about the privacy of individuals' personal information and because many jurisdictions are creating new legislative and regulatory frameworks for the conduct of research using such personal information. To supplement the views of the interview respondents, we undertook a literature review to identify practical issues faced by researchers and data custodians using population-based health and health services data for research in Canada.

Eight primary issues related to privacy protection that face researchers and data custodians were identified

from a review of academic and grey literature. These are discussed in relation to facilitating access to data while protecting individual privacy.

1. Consent - Canada has developed what has been termed a “patient-centred” model of privacy protection. Secondary uses of data in this model require special permission or special conditions before consent can be dispensed with and some mechanism is usually required to mediate competing interests. But there is considerable debate over the circumstances when explicit consent is needed for the secondary use of personal information for research purposes, leaving researchers confused about their ethical and legal obligations.

2. Data linkage - Data linkage provides for much more powerful analysis than a single data set but raises a series of concerns beyond issues of consent. There is a significant degree of variation in the access and linkage policies of data custodians among provinces. Moreover the capacity and resources to undertake the work vary considerably across the country.

3. Retention and destruction - There is no consistent approach to how data should be archived, the length of time the data should be retained, or protocols for future access, including use for audit purposes. The Canadian research community, including research funding agencies, is beginning to discuss how to better support data infrastructure and allow for compliance with emerging regulatory frameworks for the protection of privacy in research.

4. Security safeguards - There is unanimous agreement about the need for tight safeguards for the protection of data. Data custodians must be clear about what steps they take to safeguard the information they hold and must be transparent and accountable about their processes.

5. Review and oversight and the role of research ethics boards - There is consensus that some clearly

competent and independent group must review research proposals to assess the trade-offs between the risks to individual privacy and the societal benefits of the research, and to ensure all possible steps are taken to maintain confidentiality. Canada is increasingly looking towards research ethics boards to play this role, and these will need national mechanisms to ensure consistency in their work (see below).

6. Multiple rules, policies and procedures - Multiple rules, policies and procedures, which vary across jurisdictions and organizations, govern access to research data in Canada. The lack of standardization in access procedures, and in data quality, extraction and linkage, are immensely frustrating for researchers, especially those who wish to work with multiple data sets (and thus custodians), or those who wish to engage in cross-jurisdictional work.

7. Public communication - Engaging the public in discussions about research, privacy and use of data was seen as very important. A view held by many in the research community is that the general public does not understand the importance of the research being undertaken or its potential societal benefits and therefore needs to be convinced of its utility. At least two projects underway in Canada aim to assess public views on the use of personal information in research, and to develop better public communication tools.

8. Legal and policy frameworks surrounding data access and privacy protection - The existing regulatory framework in Canada exhibits clear policy support for non-consensual use of personal information for research purposes, but there is considerable variation in the practicalities of doing so. The federal *Personal Information Protection and Electronic Documents Act (PIPEDA)* and various new provincial privacy laws have added to the complexity of the legislative environment in Canada, and introduced ambiguities about the necessary

steps needed for researchers to comply with privacy legislation, especially for cross-provincial work. CIHR's privacy best practices guidelines are one step toward harmonization or consistency, but it remains to be seen how these guidelines will work with or influence changes in privacy laws across Canada or in much needed changes from a population health and health services research perspective in the existing Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans.

A number of potential options for improving access to data for health research while strengthening privacy safeguards were identified in this review. Some of these are already underway in Canada.

1. Develop privacy tool kits - A tool kit that research organizations and data custodians could use to protect privacy while allowing access to data could help to standardize practices. Appropriate tool kits would include techniques for masked (i.e. anonymized) data sharing, techniques for secure transfer of data, consent forms, and procedures to reduce re-identification.

2. Develop best practice privacy guidelines or standards - Because it is the interpretation of legislation that shapes approaches to the use and access of data, the development of best practices guidelines or voluntary standards for protecting privacy and confidentiality can support harmonization. A pan-Canadian Health Information Privacy and Confidentiality Framework has been developed by federal, provincial and territorial government officials with a view to creating a harmonized series of legislative provisions to protect personal health information.

3. Develop models of data stewardship - The roles of data stewards are worth clarifying and developing further. Separating those holding the data from those undertaking the research by relying on independent assessments of privacy risks and confidentiality protocols removes any actual or perceived conflict of

interest. Success would depend on a credible process, quick turnaround, transparent decision-making and assurance of some form of meaningfully independent oversight and accountability.

4. Strengthen and improve the practices of research ethics boards - Further work is currently underway to identify innovative best practices and delineate variation in policies and practices of research ethics boards in governing privacy, confidentiality and security issues in health research.

5. Public communication about research and privacy trade-offs - Public discussion about balancing research and privacy issues could be facilitated through targeted information about data collection and use, templates for effective communication, and relationship-building with reporters interested in health issues.

Existing inventory and data access activities

An in-depth analysis of relevant initiatives to enhance documentation of and access to data resources revealed significant efforts, international and Canadian, to build inventories of data. In the US and the UK, this work involves the development of data archives, inventories of databases, and web-based portals aimed at researchers, as well as a number of complementary activities. There are a broad variety of approaches and some evolving best practices, and a complex and evolving scientific agenda related specifically to the documentation of data and research resources. Most efforts so far concentrate on the documentation of survey data, often in combination with providing actual access to those data. No best practice models were specifically targeted to the areas of population health and health services research, which cover a vast set of content issues and rely on a similarly wide and increasingly complex variety of data sources.

In Canada, most inventory-building activity is organization-specific and aimed at documenting

organizational data holdings only. There is little standardization in approaches and it is not clear how useful these inventories really are to the research community. There is also little coordination in improving data documentation and access. There is no single Canadian portal to identify data sources, no standard format being used to compile information, the sources are in varying states of maintenance, there is spotty coverage by agency, and hence only narrow topic-specific information of variable quality. Finally, there is little sustained effort to provide such access in the area of administrative data, an important resource for population health and health services researchers. In short, there is no coordinated and focused development that could provide a strong foundation for Canada's research community.

Building an inventory of databases

We were initially given the task of developing an electronic framework for the creation of an inventory of databases relevant to population health and health services research in Canada. We were able to develop a framework that might serve as the content infrastructure for such an endeavour, but our review of the international and Canadian inventory-building activities suggests that actually building a prototype “inventory” would be quite premature.

Three major areas of consideration will need to be addressed by organizations wishing to develop such an inventory. The first is the *model*—what is the nature of the inventory, how often will it be updated, and so on; the second is *stewardship and management*—who will assume responsibility for building, populating and maintaining such an inventory?; and the third is *funding*—from what source(s) will the considerable funding needed for both start-up and ongoing operating costs be derived? In addition to these issues, potential funders of an inventory must also consider how this effort fits with other work that is currently underway. International efforts in particular show the benefit of building more than a basic inventory of data sets. Additional efforts to preserve investments in research data and

ultimately enhance our understanding of health and the factors that determine it are required.

We did, however, develop and pilot test a first version of a data collection tool that can provide the basis for creating a population health and health services research inventory of databases. A number of existing resources were reviewed to identify the best ways to collect information about databases, including general descriptive information, attributes of the data such as the unit of observation and availability of the data for research. From these, we developed a conceptual model to support consistent recording of information about the content of data sets in a manner that provides relevant information about the population and public health and health services research landscape. The utility of the tool was tested, and a prototype ‘database’ was developed from a diverse sample of candidate data sets. We also created a decision tool to be used for deciding whether a particular database is relevant for inclusion in the inventory.

Next steps and recommendations

Canada has an international reputation based on the development and implementation of a population health framework—an understanding and recognition of the many factors that influence the health status of individuals and populations. Canada is also known for the collection and research use of administrative data related to health care services. In part, this reputation is based on the availability of universal and comprehensive data about the use of health care services, data that exist because of the funding and administrative structures of provincial, territorial and federal health care services. This reputation also comes from recognition of Canadian researchers as innovators in understanding the power that such data hold, and in converting that understanding into research findings that have provided a wealth of evidence for the policy development process.

Our work suggests, however, that Canada is not currently recognized as a leader when it comes to the

systematic organization, archiving, documentation of, and access to data relevant to population health and health services research. Our ten recommendations highlight opportunities to change the situation.

Recommendation 1:

CIHR should take a lead in coordinating a series of activities to address privacy issues that are specific to the population health and health services research community. This work includes:

- a) Clarifying the definition of research that has “public value”;
- b) Developing a constellation of privacy tools and techniques (including best practice guidelines) to assist researchers and data custodians in protecting privacy while allowing access to data;
- c) Strengthening the role of research ethics boards, increasing and harmonizing expertise;
- d) Influencing the development and interpretation of regulatory and legislative frameworks to ensure they support privacy-sensitive research, and where possible, that they are harmonized across jurisdictions; and
- e) Engaging with the public about the value of health and health services research and how it should be conducted in a privacy-sensitive manner.

Recommendation 2:

CIHR should convene and lead a “coordinating body” that will focus on improving access to population health and health services research data and that will be charged with reviewing and carrying forward the recommendations in this report.

Recommendation 3:

CIHR, as the lead organization for health research in the country, and in cooperation with other funders of health research, should strongly encourage key national and provincial data custodians to review their mandates, with the goal

of clarifying and increasing their commitment to providing data and other supports for population health and health services research.

Recommendation 4:

Data custodians of population health and health services data, including the Canadian Institute for Health Information and provincial data custodians, should be encouraged to work with privacy experts and the research community to create and make available public use microdata sets as well as to provide access to more detailed microdata sets for publicly funded research.

Recommendation 5:

Provincial and regional custodians of population health and health services data should develop clear processes and equitable costing mechanisms for making data available to researchers.

Recommendation 6:

CIHR should support the costs of conducting data-based research in population health and health services research by:

- 1) Under certain circumstances, allowing operational research budgets to include the costs of archiving and documenting large-scale data collection efforts, where there is intent to make those data more broadly available to the research community;
- 2) Developing funding streams that parallel the “equipment grant” program used by the basic and clinical health research domains.

Recommendation 7:

CIHR should actively pursue opportunities to work with current initiatives with the potential to improve access to research data that supports development of population health and health services research:

- a) In the ongoing National Consultation on Access to Scientific Research Data, to ensure that the special circumstances around access to population health

and health services data (i.e. privacy considerations around personal health information and dependence on non-research data collectors and custodians) are addressed;

- b) In influencing Canada Health Infoway, to explicitly consider and build in mechanisms to support researcher access to data as it invests in prototypical development of information infrastructure.

Recommendation 8:

CIHR should work with partners to develop a web-based “population and public health and health services research” portal that could house an electronic inventory as well as related tools to support the research community to use existing data resources efficiently and in a privacy-sensitive manner.

Recommendation 9:

The partners should review the findings from the interviews and the survey of existing activities to reassess their commitment to building, maintaining and refining an inventory.

Recommendation 10:

If the partners wish to proceed with development of an inventory, they should develop an appropriate vision and business plan. This vision/business plan should:

- a) define the objectives of the resource;
- b) identify the primary customers to be served;
- c) identify a model that can build on Canadian activities already underway to document agency- and topic-specific data holdings;
- d) identify a steward or host agency that can competently develop and manage the resource;
- e) identify ongoing funding to support development over a period of at least five years; and

- f) identify an evaluative process to ensure the resource developed meets the needs of all relevant stakeholders

There is a great deal that could be done to support the community of population and public health and health services researchers in Canada. Building an inventory of population-based databases, as envisioned by the funders of the RFP for this project, is one option. But there are many issues to consider prior to starting down that particular path.

There is a clear role for a body, working group, or some other organization to take the recommendations in this report and coordinate or monitor activities relevant to them. Less clear is how such a group might be formed and maintained. Our hope is that CIHR will recognize the critical importance of this work in supporting its researchers and will take on this daunting but important challenge.

Résumé

À l'automne 2002, l'Institut de santé publique et des populations et l'Institut des services et des politiques de la santé des IRSC ont émis une demande de propositions conjointement avec l'Initiative sur la santé de la population canadienne (laquelle fait partie de l'ICIS), le Centre de coordination de la surveillance de Santé Canada et Statistique Canada. Cette demande de propositions visait à définir l'état actuel des bases de données sur la santé des populations et les services de santé au Canada ainsi qu'à démontrer combien il est possible de les utiliser pour faire des recherches innovatrices et importantes en santé. Aux fins du présent projet, les données sur la santé des populations et les services de santé ont été désignées comme des bases de données administratives, des banques de données d'enquêtes et des registres qui sont représentatifs de toute une population qui habite dans une région géographique donnée. Bien que le Canada dispose de quelques-uns des répertoires de données les mieux élaborés permettant l'étude de la santé et des soins de santé, dans la demande de propositions, on note que : « le défi consiste maintenant à améliorer l'accessibilité et l'utilisation de l'infrastructure de données actuelle pour les besoins d'importants travaux de recherche en santé en vue de permettre de judicieux investissements destinés à accroître les données et la capacité d'analyse ».

Ce sont les Réseaux canadiens de recherche en politiques publiques (RCRPP), en partenariat avec le CHSPR, situé à l'Université de la Colombie-Britannique, qui ont reçu le mandat d'exécuter ce projet. Le présent rapport : montre les résultats des entrevues auxquelles ont participé les personnes chargées de collecter des données et de les garder ainsi que les utilisateurs pour définir les enjeux actuels et nouveaux associés à la collecte, au stockage et à l'utilisation des données; examine comment se dessinent les enjeux actuels associés à l'accès à l'information et à la protection de la vie privée au Canada; étudie les activités qui ont lieu au Canada et à l'échelle internationale visant à fournir

de l'information sur des ensembles de données et leur accès; présente les raisons qui motivent la création d'un inventaire de bases de données sur la santé des populations et les services de santé; décrit le prototype d'un outil servant à recueillir des données, lequel aiderait à mettre au point un tel inventaire; présente des recommandations visant à promouvoir l'amélioration de l'accès aux données canadiennes et de leur utilisation dans les secteurs de recherche sur la santé des populations et les services de santé.

En raison de l'ampleur du travail, nous avons été en mesure de formuler et de présenter des recommandations sur des sujets majeurs, mais nous avons été incapables d'aborder en détail chacune des préoccupations. Au fur et à mesure que l'étude progressait, le contexte du projet a évolué à une telle vitesse que nous avons été incapables de traiter de tout ce qui pouvait sembler pertinent ou intéressant. Même s'il nous a été impossible de suggérer des solutions immédiates dans tous les cas ou de couvrir entièrement l'ensemble des sujets, nous espérons que le présent rapport servira à orienter les mesures qui pourraient être prises pour appuyer les chercheurs qui étudient la santé des populations et les services de santé au Canada.

Contexte

Au Canada, de nombreux intervenants ayant tous des mandats et des rôles distincts participent à la collecte de données. Il existe bien des dépositaires de données, tant à l'échelle nationale, provinciale que territoriale, qui recueillent et conservent un large éventail de données sur la santé des populations et les services de santé.

Par contre, ces collecteurs et dépositaires de données, qui incluent notamment Statistique Canada, l'ICIS, Santé Canada ainsi que les ministères provinciaux, territoriaux et fédéraux, n'ont généralement pas le mandat explicite d'appuyer le milieu de la recherche. Les données sont collectées à des fins de santé

publique et de surveillance, ou plus fréquemment dans le cadre du fonctionnement des systèmes de santé et d'éducation ou des systèmes sociaux.

Indépendamment de la raison qui motive la collecte de données, une seconde analyse de celles-ci nous permettrait grandement d'approfondir notre compréhension des effets qu'ont les politiques publiques et d'autres interventions sur les particuliers et les populations. Cependant, une seconde analyse requiert aussi la mise en place d'une série de dispositions complexes visant à régir la conservation, la divulgation et l'utilisation des données. Ces opérations peuvent être particulièrement contestées lorsque la recherche ne fait pas partie du mandat premier de l'organisme qui est chargé de collecter les données.

En conséquence, le Canada a mis en place une série de dispositions qui est complexe et parcellisée permettant à certains chercheurs seulement d'avoir accès à des données. Les entrevues que nous avons effectuées auprès des utilisateurs, collecteurs et dépositaires de données nous ont permis de mettre l'accent sur ces questions.

Points de vue des utilisateurs, collecteurs et dépositaires de données

Quarante-trois personnes de partout au pays ont participé à l'entrevue et répondu aux questions suivantes : définir les enjeux qui les préoccupent le plus; voir si un inventaire électronique des bases de données sur la santé des populations et les services de santé serait pour eux un outil utile; établir les secteurs d'investissement futurs.

Accès

Nous avons rapidement constaté que chez les utilisateurs de données, c'est l'accès qui constitue la principale préoccupation tandis que chez les collecteurs et dépositaires de données, c'est de répondre aux demandes d'accès des chercheurs.

Au moment d'organiser les observations et les suggestions que nous avons reçues des participants lors des entrevues, nous avons eu recours à un cadre bien établi qui décrit plusieurs aspects de l'accès, dont chacun reflète les enjeux et les défis auxquels se heurtent les utilisateurs, collecteurs et dépositaires de données ainsi que les possibilités d'élaboration, de maintenance et d'utilisation de bases de données sur la santé des populations et les services de santé.

1. **Disponibilité.** Les organismes de recherche devraient faciliter les discussions entre chercheurs/utilisateurs et collecteurs/dépositaires pour améliorer l'accès aux données. De plus, des ressources devraient être allouées pour promouvoir l'interconnexion des bases de données et la création de liens entre les autorités gouvernementales.
2. **Accessibilité.** Les utilisateurs ont suggéré que chaque province se dote d'un programme de centres de données de recherche de Statistique Canada et ont d'ailleurs recommandé la création d'un plus grand nombre de ces centres dans chacune des provinces. Certains des dépositaires étudient actuellement des façons d'étendre l'accès aux données en assouplissant les politiques d'accès ou en renégociant des contrats de licence.
3. **Organisation.** Certains utilisateurs demandent qu'on définisse une vision pancanadienne de l'accès aux données, qu'on élabore une politique uniforme et normalisée en ce sens et qu'on mette de côté les multiples arrangements ponctuels. D'autres proposent la création d'un seul registre de données dans chacune des provinces dont l'objectif unique serait d'assurer l'accès à des données de qualité provenant de tous les secteurs gouvernementaux. D'autres encore suggèrent la mise au point d'un format universel qui servirait à exporter les données et à en fournir de meilleures pour que les utilisateurs puissent les comprendre

et les utiliser plus facilement. Les collecteurs et dépositaires de données recommandent notamment l'assouplissement des contrats de licence conclus avec les centres qui reçoivent des données du gouvernement provincial. La plupart d'entre eux ont suggéré l'augmentation du financement pour tenir compte particulièrement du fait que les données qu'ils recueillent font partie de leurs activités de base.

4. **Accessibilité économique.** À l'exemple des États-Unis, les utilisateurs recommandent la gratuité des données publiques et leur accès par l'entremise d'Internet (à condition de mettre en place des mesures de contrôle qui assurent adéquatement la protection de la vie privée). Les collecteurs et dépositaires de données suggèrent l'attribution d'un financement pluriannuel et la reconnaissance de la nécessité de financer les coûts associés à la maintenance et au nettoyage des données. Les technologies telles que le dossier médical électronique ont été perçues comme des innovations susceptibles de réduire le travail des techniciens de données responsables de mettre à jour et de nettoyer les dossiers.

5. **Acceptabilité.** Les utilisateurs recommandent un programme de sensibilisation pour traiter des préoccupations soulevées par le public à propos de la protection de la vie privée et pour démontrer la valeur de la recherche, afin d'améliorer la santé globale des populations ainsi que le fonctionnement du système de soins de santé. Certains croient que cette mesure aiderait le public à comprendre si la législation relative à la protection de la vie privée fait la distinction entre des chercheurs de bonne foi et d'autres utilisateurs de données tels les entreprises commerciales. D'autres proposent que les IRSC facilitent les discussions avec les collecteurs et dépositaires de données pour que ceux-ci puissent s'entendre sur un accès plus souple tout en garantissant la protection de la vie

privée. De manière générale, tant les utilisateurs que les collecteurs et les dépositaires ont indiqué le besoin de communiquer plus efficacement les uns avec les autres, et d'établir des rapports et un lien de confiance entre eux.

6. **Pertinence.** Les utilisateurs ainsi que les collecteurs et dépositaires de données recommandent que la qualité des données fasse partie des priorités et que les ressources nécessaires soient allouées en conséquence. La normalisation des définitions des données et des méthodes de collecte, une formation accrue du personnel, la sensibilisation des utilisateurs de données, une meilleure documentation de données et l'utilisation des technologies telles que les enquêtes assistées par téléphone et les dossiers médicaux électroniques contribueraient grandement à améliorer la qualité des données.

Inventaire électronique

Lorsqu'elle a répondu à la question sur l'utilité d'un inventaire électronique des bases de données sur les populations visant à appuyer la santé des populations et les services de santé, la majorité des utilisateurs, collecteurs et dépositaires de données ont fourni du soutien ou l'aide de leur personnel qualifié pour mettre au point une telle ressource. Si un inventaire était élaboré, bien des répondants ont suggéré qu'il soit convivial, diffusé en ligne dans un site Web dédié, interrogeable, indexé et doté de variables normalisées. De nombreux répondants ont estimé qu'il serait utile d'établir des liens vers le site Web du dépositaire, des dictionnaires de données et de la documentation de données, ainsi que des liens vers des articles ou des rapports qui se sont appuyés sur un ensemble de données. Bien des répondants ont indiqué que l'inventaire devrait être diffusé dans un portail Web pour améliorer l'accès aux données actuelles.

La plupart estiment que le dépositaire de l'inventaire devrait être un organisme national et indépendant. Il pourrait s'agir d'une nouvelle entité qui n'aurait

qu'un seul mandat et qui serait mis sur pied grâce à un accord fédéral, provincial ou territorial, ou bien d'une entité nationale existante telle que l'ICIS, l'Institut des services et des politiques de la santé et l'Institut de la santé publique et des populations des IRSC, ou Santé Canada.

Engagements recommandés

Les répondants avaient de nombreuses suggestions à faire pour créer de nouveaux ensembles de données, notamment ce qui suit :

- des données sur les services de santé – soins communautaires, santé mentale, santé publique, toxicomanie;
- des données sur la santé des populations – maladies chroniques, évolution des maladies;
- des données sur le mesurage biologique et physique;
- des données découlant d'études longitudinales – aînés, enfants, études de cohortes;
- des données sur des populations particulières – Autochtones, sans-abri.

D'autres ont proposé que des engagements soient pris au sujet de ce qui suit :

- négocier des politiques de protection de la vie privée et l'accès aux données normalisés;
- assurer un financement stable et continu pour maintenir les bases de données et en acheter;
- normaliser les ensembles de données existants et créer des documents de données;
- former davantage les chercheurs sur l'utilisation de grands ensembles de données et l'exécution d'analyses secondaires ainsi que les techniciens pour qu'ils puissent soutenir et maintenir de grands ensembles de données;
- faciliter les comparaisons interrégionales et l'interconnexion des données surtout en ce qui a trait à celles sur les services de santé et aux déterminants de la santé;
- voir à ce que les collecteurs/dépositaires et les utilisateurs/chercheurs apprennent à mieux se connaître;

- définir une vision et une stratégie nationales visant la collecte, la maintenance et la mise en commun de données qui sont financées par les fonds publics du Canada.

Enjeux associés à la protection de la vie privée et à l'accès

Le contexte de la recherche est de plus en plus complexe en raison des préoccupations que soulève le public sur la protection des renseignements personnels des particuliers et des nombreuses autorités gouvernementales qui élaborent de nouveaux cadres législatifs et réglementaires pour régir les recherches où on utilise de tels renseignements. Pour appuyer les points de vue des répondants, nous avons entrepris d'analyser des documents afin de cerner les enjeux dont doivent traiter les chercheurs et dépositaires de données qui utilisent les données sur la santé des populations et les services de santé pour faire des recherches au Canada.

En passant en revue des publications universitaires et de la littérature grise, nous avons établi que les chercheurs et les dépositaires de données sont aux prises avec huit difficultés qui sont liées à la protection de la vie privée. Celles-ci sont présentées ci-dessous, en tenant compte des mesures à prendre pour faciliter l'accès aux données et protéger les renseignements personnels.

1. Consentement. Le Canada a mis au point un modèle de protection de la vie privée que l'on dit *taxé sur le patient*. Consentir à l'utilisation secondaire de données selon ce modèle nécessite une permission spéciale ou l'existence de conditions particulières, et un mécanisme quelconque est habituellement requis pour concilier des intérêts qui sont concurrents. Toutefois, un débat important s'est engagé, lequel tourne autour des circonstances réclamant que l'utilisation secondaire de renseignements personnels à des fins de recherche soit formellement autorisée, semant ainsi la confusion parmi les chercheurs quant à leurs obligations morales et légales.

2. **Interconnexion des données.** Le couplage de données, par comparaison à l'utilisation d'un seul ensemble de données, permet de faire des analyses qui sont beaucoup plus exhaustives, mais ce procédé soulève une série de préoccupations qui dépassent la question du consentement. Les politiques d'accès aux données et d'interconnexion à celles-ci, auxquelles sont assujetties les dépositaires, diffèrent considérablement d'une province à une autre. En outre, les capacités et les ressources requises pour entreprendre un tel travail varient de manière importante partout au pays.
3. **Conservation et élimination des données.** Il n'existe pas de méthode cohérente pour définir la façon d'archiver les données et leur période de conservation, ou de protocoles établis pour déterminer leur utilisation future, notamment leur utilisation à des fins de vérification. Le milieu de la recherche du Canada, particulièrement les organismes de financement de la recherche, a commencé à discuter de la façon de soutenir plus efficacement l'infrastructure des données et d'assurer la conformité aux nouveaux cadres réglementaires visant à protéger la vie privée dans la recherche.
4. **Dispositifs de protection.** On s'entend unanimement sur le besoin de mettre en place des dispositifs de sécurité à des fins de protection des données. Les dépositaires de données doivent définir clairement les étapes à franchir pour protéger les renseignements qu'ils protègent, faire preuve de transparence dans leurs processus et rendre des comptes en ce sens.
5. **Examen, surveillance et rôle des comités d'éthique de la recherche.** On s'accorde à dire qu'un groupe qui est tout à fait compétent et indépendant doit examiner les propositions de recherche pour évaluer les risques de divulgation de renseignements personnels par rapport aux avantages sociétaux de la recherche, et de voir à ce que toutes les mesures nécessaires soient prises pour respecter la confidentialité de ces renseignements. Le Canada se tourne de plus en plus vers des comités d'éthique de la recherche pour jouer ce rôle. Ceux-ci devront se doter de mécanismes nationaux pour assurer la cohérence dans leur travail (voir ci-dessous).
6. **Règles, politiques et procédures multiples.** Il existe de multiples règles, politiques et procédures, lesquelles diffèrent d'une autorité gouvernementale et d'une organisation à une autre, qui régissent l'accès aux données de la recherche au Canada. Le manque de normalisation en matière de procédures d'accès, de qualité des données, d'extraction et d'interconnexion engendre de grandes frustrations parmi les chercheurs, surtout parmi ceux qui souhaitent travailler avec de nombreux ensembles de données (par conséquent, les dépositaires) ou ceux qui souhaitent entreprendre un travail d'envergure intergouvernemental.
7. **Communication avec le public.** La participation du public aux discussions sur la recherche, la protection de la vie privée et l'utilisation des données sont perçues comme des engagements importants. Bon nombre de personnes du milieu de la recherche estiment que le grand public ne comprend pas l'importance de la recherche qui a été entreprise ou des avantages sociétaux possibles, et ont donc besoin d'être convaincues de son utilité. Il y a au moins deux projets qui sont en cours au Canada qui visent à étudier les points de vue du public sur l'utilisation des renseignements personnels à des fins de recherche et à mettre au point de meilleurs outils de communication avec le public.
8. **Cadres juridique et politique relatifs à l'accès aux données et à la protection de la vie privée.** Le cadre réglementaire existant au Canada prévoit

un soutien à la politique relative à l'utilisation non autorisée de renseignements personnels à des fins de recherche, mais les considérations pratiques en ce sens diffèrent énormément. La *Loi sur la protection des renseignements personnels et des documents électroniques* (LPRPDE) du gouvernement fédéral et diverses autres nouvelles lois provinciales relatives à la protection de la vie privée ont contribué à complexifier l'environnement législatif au Canada et rendent ambiguës les étapes essentielles que doivent suivre les chercheurs pour respecter les lois en matière de protection de la vie privée, notamment en ce qui a trait au travail qui se fait à l'échelle interprovinciale. Les lignes directrices des IRSC qui concernent les pratiques exemplaires en matière de protection de la vie privée sont à peu près harmonisées ou cohérentes, mais reste à savoir comment elles seront appliquées, pourront influencer, à l'échelle du Canada, les changements aux droits relatifs à la protection de la vie privée ou engendrer des changements qui sont tout à fait nécessaires à l'*Énoncé de politique des trois Conseils : Éthique de la recherche avec des êtres humains*, dans une perspective de recherche sur la santé des populations et les services de santé.

Dans le cadre de cet examen, un certain nombre d'options possibles ont été définies pour améliorer l'accès aux données à des fins de recherche en santé tout en renforçant les dispositions relatives à la protection de la vie privée. Certaines de ces options ont déjà été mises en œuvre au Canada.

1. Mise au point d'une trousse d'information sur la protection de la vie privée. Il s'agit d'une trousse que les organismes de recherche et les dépositaires de données pourraient utiliser pour protéger la vie privée tout en permettant l'accès aux données, qui aiderait à normaliser les pratiques utilisées. Une trousse d'information appropriée inclurait des techniques servant à

masquer (rendre anonyme) la mise en commun des données et à transférer de manière sécuritaire ces dernières et des formulaires de consentement ainsi que des procédures visant à réduire la répétition des identifications.

- 2. Élaborer des lignes directrices ou des normes de pratiques exemplaires en matière de protection de la vie privée.** Puisque c'est l'interprétation de la loi qui détermine la façon d'utiliser les données et d'accéder à celles-ci, l'élaboration de lignes directrices sur les pratiques exemplaires ou de normes volontaires visant à protéger la vie privée peuvent favoriser l'harmonisation. Un cadre pancanadien de protection de renseignements médicaux personnels et de la vie privée a été élaboré par des représentants des gouvernements fédéral, provinciaux et territoriaux en vue de mettre en place une série de dispositions législatives harmonisées pour protéger les renseignements médicaux personnels.
- 3. Créer des modèles de gérance des données.** Il importe de clarifier et de définir davantage les rôles des gestionnaires de données. Faire la distinction entre ceux qui ont la charge de garder les données et ceux qui font des recherches en s'appuyant sur des évaluations indépendantes des risques sur la vie privée et des protocoles sur la confidentialité élimine tout conflit d'intérêts, qu'il soit réel ou perçu. Le succès du modèle se fonde sur un processus crédible, un temps de réponse rapide, un processus décisionnel transparent et la garantie qu'il y ait en place une certaine forme de processus de surveillance et de reddition de comptes autonome.
- 4. Renforcer et améliorer les pratiques des comités d'éthique de la recherche.** D'autres travaux sont actuellement en cours pour établir des pratiques exemplaires innovatrices et définir les divergences existantes dans les politiques et les pratiques des comités d'éthique de la recherche

qui servent à gérer les questions de protection de la vie privée et de sécurité dans la recherche en santé.

5. *Communication avec le public sur l'harmonisation de la recherche et la protection de la vie privée.* Il serait possible de faciliter une discussion publique sur les enjeux liés à l'harmonisation de la recherche et de la protection de la vie privée en ciblant l'information sur la collecte et l'utilisation des données, en ayant recours à des modèles de communication efficaces et en développant des relations avec les journalistes qui s'intéressent aux questions relatives à la santé.

Inventaire existant et activités liées à l'accès aux données

Une analyse approfondie d'initiatives pertinentes visant à enrichir la documentation des ressources sur les données et à élargir l'accès à ces dernières a révélé qu'au Canada et à l'étranger des efforts importants sont déployés pour bâtir des inventaires de données. Aux États-Unis et en Grande-Bretagne, les travaux amorcés en ce sens comprennent l'élaboration d'archives de données, d'inventaires de bases de données et des portails Web qui sont destinés aux chercheurs, ainsi que la tenue d'un certain nombre d'activités complémentaires. Il existe un large éventail de démarches, quelques pratiques dynamiques et un programme scientifique qui est complexe et en évolution portant précisément sur la documentation des données et les ressources en matière de recherche. La plupart des efforts fournis à ce jour ont été centrés sur la documentation de données d'enquête qui souvent, par la même occasion, étaient rendues accessibles. Aucuns modèles sur les pratiques exemplaires en particulier n'étaient visés dans les domaines de recherche sur la santé des populations et les services de santé, lesquels englobent un large ensemble d'enjeux sur le contenu et reposent sur une variété de sources de données aussi vaste et de plus en plus complexe.

Au Canada, la majorité des activités d'élaboration d'inventaires ont lieu à l'échelle de l'organisme seulement et ont pour objectif de documenter ses propres fonds de données. Les démarches utilisées sont peu normalisées, et il est difficile de déterminer combien utiles en fait sont ces inventaires pour le milieu de la recherche. De plus, très peu d'efforts sont fournis pour coordonner l'amélioration de la documentation des données et de l'accès à celles-ci. Il n'existe aucun portail canadien proposant des sources de données; aucun format normalisé n'est utilisé pour compiler l'information; la maintenance des sources de données n'est pas uniforme, et les organismes stockent leurs données de façon irrégulière, ce qui se traduit par la diffusion d'information qui est axée sur un nombre de thèmes limité et qui est de qualité inégale. Enfin, très peu d'efforts sont fournis pour donner accès à des données administratives, une ressource importante pour les chercheurs qui travaillent dans le domaine de la santé des populations et les services de santé. En résumé, il n'existe pas de développement coordonné et ciblé d'inventaires de bases de données qui fournirait au milieu de la recherche du Canada une solide assise.

Élaboration d'un inventaire de bases de données

Notre tâche consistait initialement à élaborer un cadre électronique visant à créer un inventaire de bases de données relatives à la recherche sur la santé des populations et les services de santé au Canada. Nous avons été en mesure de mettre au point un cadre qui pourrait servir d'infrastructure de contenu à un tel projet, mais l'examen que nous avons fait des activités qui sont menées à l'échelle canadienne et internationale pour concevoir des inventaires révèle qu'il serait tout à fait prématuré de bâtir un « prototype » d'inventaire.

Les organismes désireux de bâtir un tel inventaire auront à prendre en considération les trois importants aspects suivants. Il s'agit d'abord de définir le *modèle* pour en déterminer sa nature, la fréquence

à laquelle il sera mis à jour, etc. Ensuite, il faut considérer la *gérance* et la *gestion* pour établir qui aura la responsabilité de bâtir un tel inventaire, de l'entretenir et de le tenir à jour. Enfin, il y a la question du *financement* : qui sera le principal bailleur de fonds (ou les principaux bailleurs de fonds) qui fournira l'argent nécessaire pour payer les coûts de lancement et de maintien des opérations? En plus de ces questions, les commanditaires potentiels d'un inventaire doivent tenir compte de comment cette initiative cadre avec les autres travaux qui sont en cours. En évaluant ce qui se fait sur le plan international en matière d'élaboration d'inventaires, nous avons remarqué notamment qu'il est avantageux de bâtir non seulement un inventaire de base d'ensemble de données, mais aussi souhaitable de fournir d'autres efforts pour conserver les investissements qui ont été réalisés en matière de données de recherche, données qui, en fin de compte, approfondissent nos connaissances de la santé et des facteurs qui la définissent.

Nous avons cependant mis au point une première version d'un outil de collecte de données et en avons fait un essai pilote. Cet outil peut servir de fondement pour bâtir un inventaire de bases de données pour la recherche sur la santé des populations et les services de santé. Un certain nombre de ressources existantes ont fait l'objet d'un examen pour déterminer les meilleures façons de collecter l'information sur les bases de données, notamment l'information descriptive générale, les attributs de données tels que l'unité d'observation et l'accessibilité aux données de la recherche. C'est à partir de ces ressources que nous avons mis au point un modèle conceptuel destiné à enregistrer uniformément de l'information sur le contenu des ensembles de données de manière à fournir des renseignements qui sont pertinents à propos de l'aménagement de la recherche sur la santé des populations et les services de santé. L'utilité de l'outil a été mise à l'épreuve, et un prototype de « base de données » a été construit à partir d'un échantillon diversifié d'ensembles de données candidats. Nous

avons aussi mis au point un outil d'aide à la décision qui servira à déterminer s'il convient ou non d'inclure une base de données en particulier dans l'inventaire.

Prochaines étapes et recommandations

Le Canada est reconnu à l'échelle internationale pour avoir élaboré et mis en œuvre un cadre sur la santé des populations, puisqu'il comprend et reconnaît les nombreux facteurs qui influent sur l'état de santé des particuliers et des populations. Il est aussi reconnu pour sa capacité de recueillir et d'utiliser à des fins de recherche des données administratives sur les services de soins de santé. Sa réputation est partiellement attribuable au fait que des données universelles et exhaustives sur l'utilisation des services de santé sont disponibles, en raison des structures de financement et d'administration ayant été établies au sein des services de soins de santé provinciaux, territoriaux et fédéraux. Elle est aussi imputable à la renommée des chercheurs canadiens, qui sont considérés comme des innovateurs puisqu'ils comprennent la valeur de ces données et sont en mesure de s'en servir pour en tirer des conclusions. Celles-ci constituent une abondance d'informations utiles au processus d'élaboration de politiques.

Toutefois, le travail que nous avons accompli révèle que le Canada n'est actuellement pas reconnu à titre de chef de file en termes d'organisation systématique, d'archivage et de documentation des données qui sont utiles à la recherche sur la santé des populations et les services de santé, et en ce qui a trait à l'accès à ces données. Nous avons formulé dix recommandations qui permettraient de changer la situation.

Recommandation 1 :

Les IRSC devraient prendre l'initiative de coordonner une série d'activités pour traiter des enjeux qui sont associés à la protection de la vie privée et qui sont propres au milieu de la recherche sur la santé des populations et les services de santé. Ce travail comprend ce qui suit :

- a) éclaircir ce que veut dire une recherche qui est « utile pour le public »;
- b) mettre au point une panoplie d'outils et de techniques pour protéger la vie privée (notamment des lignes directrices sur les pratiques exemplaires) afin d'aider les chercheurs et dépositaires de données à protéger la confidentialité tout en permettant l'accès aux données;
- c) renforcer le rôle des comités d'éthique en recherche, accroître et harmoniser le savoir-faire;
- d) exercer des pressions pour que des cadres législatifs et réglementaires soient mis au point, qui serviront à appuyer la recherche dans laquelle entre en compte la protection de la vie privée, influencer l'interprétation qu'on fait de ces cadres et voir, dans la mesure du possible, à ce qu'ils soient harmonisés dans l'ensemble des autorités gouvernementales;
- e) engager la discussion avec le public sur la valeur que la recherche sur la santé des populations et les services de santé a à leurs yeux, et comment cette recherche devrait être conduite en tenant compte de la protection de la vie privée.

Recommandation 2 :

Les IRSC devraient former et diriger un « organisme de coordination » qui veillerait surtout à améliorer l'accès aux données de la recherche sur la santé des populations et les services de santé. Il incomberait d'ailleurs à cet organisme d'examiner et d'exécuter les recommandations présentées dans le présent rapport.

Recommandation 3 :

Les IRSC, en qualité d'organisation directrice de la recherche sur la santé au pays, et en collaboration avec d'autres bailleurs de fonds de la recherche sur la santé, devraient fortement encourager les principaux dépositaires de données du Canada et des provinces à réviser leur mandat pour le préciser et s'engager davantage à fournir des données et d'autres soutiens à la recherche sur la

santé des populations et les services de santé.

Recommandation 4 :

On devrait exhorter les dépositaires de données sur la santé des populations et les services de santé, y compris l'ICIS et les dépositaires de données provinciaux, à travailler avec les spécialistes de la protection de la vie privée et le milieu de la recherche pour créer des ensembles de microdonnées et permettre leur accès au public, ainsi qu'à fournir un accès à des ensembles de microdonnées plus détaillés à des fins de recherches étant subventionnées par des fonds publics.

Recommandation 5 :

Les dépositaires provinciaux et régionaux de données sur la santé des populations et les services de santé devraient concevoir des processus qui sont clairs et des mécanismes d'établissement des coûts qui sont équitables pour rendre les données accessibles aux chercheurs.

Recommandation 6 :

Les IRSC devraient assumer les coûts associés aux recherches sur la santé des populations et les services de santé qui sont axées sur les données en faisant ce qui suit :

- 1) dans certaines circonstances, octroyer des budgets de fonctionnement pour la recherche afin de tenir compte des coûts liés à l'archivage des données et à la documentation de collections de données de grande envergure, lorsque l'intention est d'en étendre l'accès au milieu de la recherche;
- 2) trouver des volets de financement qui égalent le programme de « subvention d'achat d'appareils » utilisé dans les domaines de recherche fondamentale et clinique.

Recommandation 7 :

Les IRSC devraient chercher activement des occasions de travailler à des initiatives qui sont en cours en vue d'améliorer potentiellement

l'accès aux données de recherche, lesquelles aideraient à faire des recherches sur la santé des populations et les services de santé.

- a) Par l'entremise de la Consultation nationale sur l'accès aux données de recherche scientifique, une initiative qui est en cours, les IRSC pourraient voir à ce que les enjeux sur les circonstances particulières touchant l'accès aux données de recherche sur la santé des populations et les services de santé soient abordés (c.-à-d. les facteurs liés à la protection de renseignements médicaux personnels et la dépendance des chercheurs vis-à-vis des collecteurs et dépositaires de données ne provenant pas de la recherche).
- b) Les IRSC pourraient exercer des pressions auprès d'Inforoute Santé du Canada pour que ce dernier considère de très près des mécanismes qui faciliteraient l'accès des chercheurs aux données et intègrent ces mécanismes à ses activités au moment d'élaborer des prototypes d'infrastructure de l'information.

Recommandation 8 :

Les IRSC devraient travailler avec des partenaires pour créer un portail Web sur « la recherche sur la santé des populations et les services de santé » qui pourrait contenir un inventaire électronique et d'autres outils pertinents pour aider le milieu de la recherche à utiliser les sources de données existantes efficacement tout en respectant la confidentialité.

Recommandation 9 :

Les partenaires devraient passer en revue les résultats des entrevues et des enquêtes découlant d'activités existantes pour réévaluer leur engagement à l'égard de la création, la maintenance et l'amélioration d'inventaires.

Recommandation 10 :

Si les partenaires souhaitent aller de l'avant avec la mise sur pied d'un inventaire, ils devraient définir une vision et élaborer un plan d'activités.

Ceux-ci devraient comprendre ce qui suit :

- a) définir les objectifs de l'inventaire;
- b) identifier les principaux clients;
- c) établir un modèle qui tient compte des activités qui ont déjà lieu au Canada pour documenter les fonds de données selon les organismes et les thèmes;
- d) déterminer qui sera le gestionnaire ou l'organisme qui se chargera d'élaborer et de gérer l'inventaire;
- e) trouver du financement continu pour soutenir l'élaboration de l'inventaire pour une période d'au moins cinq ans;
- f) définir un processus d'évaluation qui permettra de vérifier si l'inventaire répond aux besoins de tous les intervenants concernés.

Bien des mesures pourraient être prises pour aider la communauté des chercheurs œuvrant dans les secteurs de la recherche sur la santé des populations et les services de santé au Canada. La création d'un inventaire de bases de données axées sur les populations, tel que l'imaginent les commanditaires de ce projet, est une option. Par contre, il faut tenir compte de bien des enjeux avant d'opter pour une voie en particulier.

Il est évident qu'un organisme, un groupe de travail ou que toute autre organisation prendra en main la mise en œuvre des recommandations présentées dans le présent rapport et qu'il coordonnera et surveillera les activités s'y rattachant. Ce qui soulève l'incertitude, c'est la façon de former et de maintenir ce groupe. Nous espérons que les IRSC prendront conscience de l'importance capitale du travail qu'ils ont à faire pour appuyer leurs chercheurs et qu'ils décideront de relever ce défi qui est redoutable mais combien important.

Chapter 1: Setting the scene

1.1 The report

In the fall of 2002, the Canadian Institutes of Health Research (CIHR) Institutes of Population and Public Health and Health Services and Policy Research jointly issued a request for proposals (RFP) with the, Canadian Population Health Initiative (a part of the Canadian Institute for Health Information), Health Canada's Centre for Surveillance Coordination and Statistics Canada. The objectives of this RFP were to describe the current status of population-based health and health services databases in Canada and to show the potential for their use in innovative and important health research. The RFP noted that while Canada has some of the best-developed data repositories for studying health and health care, "the challenge now lies in enhancing access to and use of current data infrastructure for the purposes of conducting important health research and to make wise investments to increase data and analytic capacity. However, investments to enhance data infrastructure for health research in Canada could be guided by a better understanding of current capacity and issues regarding access to and use of data across the country" (CIHR 2002).

Since 2002, there has been an ever-increasing demand by governments and other decision makers for research evidence to inform public policy formation. There is also interest from the public in more and better use of existing data to help guide policy decisions, though in all cases a desire for more research exists only insofar as it can be achieved

without threat to the privacy and confidentiality concerns of all individuals. The whole enterprise of public accountability in health care is founded on the use of survey and utilization data to provide a clear picture of what the health care system does well and where improvements are needed. The creation of a new Public Health Agency for Canada,¹ a Canadian Patient Safety Institute,² and a series of commitments to health care reform from Canada's First Ministers³ all imply a desire to analyze, understand and report to the public more about health status, the health care system and the use of health care services. There is a community of researchers anxious to advance our sophistication in achieving these objectives, but they require data that are compiled, available and understandable to researchers.

There is a major national, cross-disciplinary effort underway (as of the end of 2004) to put Canada on the path to achieving these goals. The National Consultation on Access to Scientific Research Data aims to "help Canada maximize the value received from its publicly funded natural and medical sciences research by recommending an appropriate framework and guidelines, which will facilitate open and long-term access to data coming from that research."⁴ This process is in turn consistent with the OECD Declaration on Access to Research Data from Public Funding, to which Canada became a signatory in January 2004.⁵ The original call from the CIHR has not been superseded, but must be viewed against a constantly (and quickly) evolving context.

¹Public Health Agency of Canada, "Information Backgrounder," September 2004 http://www.phac-aspc.gc.ca/media/nr-rp/phac_e.html.

²Canadian Patient Safety Institute, "Home," <http://www.cpsi-icsp.ca/>.

³Health Canada, Health Care Renewal, "A 10-year plan to strengthen health care," September 16, 2004, <http://www.hc-sc.gc.ca/english/hca2003/fimm/>.

⁴Government of Canada, National Consultation on Access to Scientific Research Data, "Home," http://ncasrd-cnadr.scitech.gc.ca/home_e.shtml.

⁵Organisation for Economic Co-operation and Development, "Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué," http://www.oecd.org/document/15/0,2340,en_2649_34487_25998799_1_1_1_1,00.html.

The Canadian Policy Research Networks (CPRN), in partnership with the Centre for Health Services and Policy Research (CHSPR) at the University of British Columbia, was awarded the contract to undertake this study. The scope of work included reviewing the current landscape of privacy and access issues in Canada; conducting interviews with data collectors, custodians and users to identify what they believe are current or emerging issues around collection, storage and use of data; a review of conceptual frameworks used for the development of inventories of population-based databases; creation of a prototype data collection tool that could assist in the development of such an inventory; and recommendations for moving forward the agenda of improving access to and use of data in the areas of population health and health services research.

Our approach to this work was to use the interviews as a touchstone for clarifying the current issues and concerns of both data custodians and the research community. We found a wide range of issues affecting individuals and organizations across Canada. The breadth of work undertaken meant that we were able to identify and offer some recommendations for major topics, but we were not able to follow every concern in great detail. The broader context mentioned above was also evolving as we conducted this study, at a speed that made it impossible for us to include everything that might be relevant or of interest. For example, the final report of the National Consultation on Access to Scientific Research Data is not yet available, nor are several other reports relevant to the area of privacy. While we are not able to suggest immediate solutions in all cases, or to provide complete coverage of all topics, we hope this report provides useful direction for actions that can be taken to support researchers in population health and health services research in Canada. The major components of the work undertaken are presented in the following chapters:

Chapter 2. Views from data users and data custodians:

This chapter describes the results of key informant

interviews from the research/data user and the data custodian communities.

Chapter 3. Privacy and access issues in the use of population-based health and health services data:

This chapter outlines the results of a literature review related to access and use of data, concerns about privacy and confidentiality, and proposals for enhancing access and use.

Chapter 4. Existing inventory and data access activities:

This chapter outlines the results of a more in-depth analysis of relevant initiatives, both international and Canadian, to enhance documentation of and access to data resources.

Chapter 5. Building an inventory of population health and health services research databases for Canada:

This chapter reviews literature on existing conceptual frameworks to classify databases, develops a framework upon which an inventory could be built, and outlines options for building and managing an inventory in the Canadian context.

Chapter 6. Next Steps and Recommendations:

This chapter outlines the project team's recommendations for future efforts and investments in the area of population-based databases.

Appendices.

Supporting material and documentation are provided in appendices to the report.

1.2 Background

For the purposes of this project, population-based health and health services data were defined as “administrative databases, registries and survey databanks that are representative of an entire population who reside in a geographic region.” At least seven general (though not mutually exclusive)

sources of data were identified for inclusion by this definition:

Person-level administrative data—e.g. provincial data on health care provision, physician services, workers' compensation, health human resources, early development instrument (EDI), criminal justice, social services

Aggregate-level administrative data—e.g. air quality, fisheries, forestry, income data

Population/public health data—e.g. disease surveillance, special populations (e.g. aboriginal, sexual orientation, etc.), immunization registries

Disease registries—e.g. cancer registries, others including cardiovascular disease, respiratory disease

Population survey data—e.g. Census, National Population Health Survey, Canadian Community Health Survey, First Nations and Inuit Longitudinal Health Survey, special waiting time research (Health Services Access Survey) done for the Health Accord by Statistics Canada, the Joint Canada/United States Survey of Health

Special purpose data holdings—e.g. CANSIM, BC Linked Health Database, Manitoba Population Health Research Data Repository

Investigator-driven research data funded by public agencies (including CIHR, Statistics Canada, Health Canada, etc.)—e.g. Aging In Manitoba study

The current data environment in Canada involves a large number of players, all with differing mandates and roles. There are a variety of data custodians in

Canada, both national and provincial/territorial, that collect and maintain these kinds of data. At the national level, the two key organizations are Statistics Canada and the Canadian Institute for Health Information (CIHI). Statistics Canada is given the authority to collect information directly from citizens under the *Statistics Act*. CIHI does not have the same authority for direct collection but instead negotiates agreements with provincial and territorial ministries of health to receive their information and compile reports. Health Canada also maintains a number of data sets pursuant to its public health and surveillance mandate although most of the information is currently for internal use only. At the provincial and territorial level, government departments or ministries maintain service data, often for payment purposes, and sometimes make this information available for research. Some have agreements with external research centres that receive and maintain the data, others keep the data in-house but grant access for specific research projects, while others do not make data accessible.

A review of the mission statements and mandates of national agencies shows clearly that enhancing researchers' access to and use of data about health and health care is *not* a primary objective of these organizations. This is in contrast to organizations such as the National Center for Health Statistics in the United States, which describes itself as “a unique public resource for health information” and prominently provides information about its privacy and data release policies, together with a variety of tools that support microdata access, on its website.⁶ Appendix A details the missions and mandates of a selection of Canadian and US organizations involved in data and research.

The general data sources listed above indicate that a great deal of data available to and used by researchers in the areas of population health and health services

⁶National Centre for Health Statistics, “About NCHS”, <http://www.cdc.gov/nchs/about.htm>.

research in Canada are collected by organizations whose primary purpose is not conducting or supporting research. Provincial ministries of education across the country, for example, collect an array of information on students, such as their age, sex, area of residence, what schools they attend, and often their scores on major provincial examinations. These data are collected in the course of operating the education system, but have potential value well beyond providing the bare operational necessities to drive allocation of funding, such as number of students per school. Education researchers are interested, for example, in understanding if there are particular school characteristics that improve educational attainment. More broadly speaking, early childhood development researchers are interested in how early life experiences—area of residence, socioeconomic status, family composition, access to day care—affect later performance in school.

Secondary analysis of data—analysis of data already collected—has great potential for improving our understanding of the impact of public policy and other interventions on the individuals and populations affected.⁷ Secondary analysis of population-based surveys, such as the Canadian Community Health Survey, provides an opportunity to investigate differences in health status and well-being for small areas within and across provinces. But secondary analysis of data (including data collected by researchers after initial research questions are addressed) necessitates a complex set of arrangements to govern its retention, disclosure and use. The issues become particularly contentious when individuals or agencies responsible for collecting data (say for the purpose of operating a public program) do not view supporting research as part of their mandate.

Researchers in Canada and elsewhere access population-based health and health services data through a number of mechanisms. These include

public release microdata files, which make anonymized individual-level data sets available to the research community. These have been less commonly used to support the health research community in Canada than other jurisdictions. A second model for providing data access involves the development of *research data archives*, in which consortiums are created to house archival data deemed to be the best in a field of inquiry and then provide access through formal protocols; this approach has also not been widely developed in Canada. Yet another model is the *research data centre*, a mechanism that has been used by Statistics Canada to allow researchers to access data within a highly secure environment subject to privacy and confidentiality protocols. A fourth model of data access that has been prominently used by the Canadian population health and health services research community can be characterized as a *data laboratory*, in which research organizations themselves maintain data holdings to support researchers who have a formal affiliation with the organization (this has been called a “fortress model” and is exemplified by the Institute for Clinical Evaluative Sciences in Ontario and the Manitoba Centre for Health Policy), or to provide data for approved research projects to researchers outside the organization with appropriate safeguards and legal protection (this has been called a “public utility model” and is exemplified by the BC Linked Health Database, which is maintained by the Centre for Health Services and Policy Research) (Black and Roos, forthcoming). A final mechanism for providing access to data relies on providing access to *tabular data* only, in which analyses that are not sensitive from a privacy perspective are made publicly available online. Statistics Canada and CIHI offer these features on their websites.

To understand the challenges for more effectively using Canadian population health and health services research data, it is important to note that there is a

⁷This is particularly true in Canada where there is a tradition of providing public programs on a universal basis.

large group of data collectors/custodians involved in collecting data who typically have no explicit mandate to support the research community. The Canadian landscape has therefore spawned a complex, fragmented set of arrangements by which some researchers can obtain access to data for research and others cannot. This report reviews the perceptions of researchers/data users and data custodians, reviews privacy and access issues, surveys international and Canadian activities, and makes recommendations about future options to address these challenges.

References

Black, C and L Roos. (forthcoming). Linking and Combining Data to Develop Statistics for Understanding the Population's Health. Hunter E, Gibson R, Friedman D (eds) in *Health Statistics in the 21st Century: Implications for Health Policy and Practice*. Oxford, UK: Oxford University Press.

CIHR: Canadian Institutes of Health Research. Institutes of Population and Public Health and Health Services and Policy Research. 2002. *Population-based health and health service data in Canada: current status and future health research potential*. Ottawa: CIHR.

Chapter 2: Views from data users and data collectors/custodians

2.1 Introduction

There are many issues facing researchers/data users and data collectors/custodians in developing and maintaining population-based health and health services databases, and in accessing, facilitating access to, and using these databases. In recent years, these issues have become increasingly complicated in Canada with the introduction of federal and provincial privacy legislation.

The research literature is focused largely on the protection of the privacy of health information (Willison 1998); informed consent for the use of health information for research purposes (Willison et al. 2003); and the role and effectiveness of research ethics boards (Kluge 1996; Ferris 2002). Granting agencies have established policies for the ethical conduct of research (Tri-Councils 1998) and have identified privacy issues as an important research topic (CIHR 2002). Professional associations have developed codes of conduct, albeit without statutory authority for enforcement (CMA 1998), and data custodians have developed policies and procedures for collection, access, storage and disclosure (CIHI 2002).

Other policy issues have not yet received the same degree of attention as privacy and ethics. From the data collector/custodian perspective, these include support for data infrastructure (funding for creation and ongoing maintenance, data linkage, technical capacity development and continuous training, and development of new technology); support for ensuring data quality (accuracy, completeness, updating); and policy issues for stewardship and ownership other than privacy and ethics issues (cost recovery, disposal, intellectual property and publication rights). From the data user perspective, only a small number of Canadian studies have documented the difficulties in accessing and using population-based databases (Tu et al. 2004; Kephart, 2002).

One of the purposes of this project was to create a better understanding of the different perspectives of data collectors/custodians and data users. Specifically, we wanted to identify the issues of greatest concern to them, their perception of emerging issues, their thoughts on the development of an electronic inventory of databases and other suggestions for future investments. To this end, we developed a survey instrument (different for each group) and identified interview respondents from across the country. Details of methodology and the survey instruments used can be found in Appendices B and C, respectively. The results of the 43 interviews conducted are summarized here. The suggestions included in this chapter are a synthesis of responses from the interviews and informed the recommendations we have outlined in Chapter 6.

2.2 Framework

Access quickly emerged as the major concern of data users, and the responses to access as the major concern of data collectors/custodians. Penchansky and Thomas (1981) developed a framework to describe users' access to health care services in terms of five dimensions—availability, accessibility, accommodation, affordability and acceptability. While this framework was originally created to enable understanding of the relationship between patients and the health care system, we have adapted it—adding a sixth dimension of “adequacy”—and used it to organize the findings from the interviews. Access in this context is used to reflect the degree of fit between the needs and interests of data users and data collectors/custodians of population-based health and health services data. The six dimensions capture the issues, challenges and opportunities facing data users and data collectors/custodians in the development, maintenance and use of population-based health and health services databases. Each dimension is defined briefly on Table 2.1, with details of the interview responses following.

Table 2.1: Dimensions of access

Dimensions of access	Penchansky and Thomas	Modified for research—user perspective	Modified for research—collector/custodian (CC) perspective
Availability	The relationship of volume and type of existing services/resources (supply) to clients' volume and types of needs (demand)	Do the required data exist? Are users aware of the existing data? Is it possible to use them for research?	Is there desire, ability and/or authority to provide data to users?
Accessibility	The relationship between the location of supply and the location of clients (takes into account client transportation resources and travel time, distance, cost)	Are data physically convenient?	What are the issues for CCs in increasing access locations? Do CCs have the resources and authority, and confidence in the end user to make data more physically convenient?
Accommodation	The relationship between the manner in which the supply resources are organized to accept clients and the clients' ability to accommodate to these factors and the clients' perception of their appropriateness	Are policies and procedures regarding access to data appropriate? Can the requirements be reasonably met in a timely fashion?	Are user desires for ready access to data compatible with the mandates and resources of CC organizations, and with their responsibilities to provide adequate oversight and protection of data?
Affordability	The relationship of prices of services to clients' ability to pay (clients' perception of worth relative to cost is an issue here)	Are data costs reasonable?	Is available funding sufficient for data collection, maintenance and support of research?
Acceptability	The relationship of clients' attitudes about providers and providers' attitudes about clients.	Do data CCs understand user needs?	Do users understand the multiple roles and responsibilities of data CCs?
Adequacy		Are data, as currently constructed, suitable for user needs?	Are available resources sufficient to support efforts to improve data quality and documentation?

2.3 Findings—Access

Data users included university researchers (13), provincial data centre scientists (1), scientists with a disease registry (1), and scientists in special purpose institutes (3). Users conducted research with a variety of population-based databases, including administrative data (e.g. hospital discharge, physician claims, drug benefits, home care, alternative payments, national health expenditures); national/provincial cross-sectional and longitudinal surveys (e.g. National Population Health Survey, Canadian Community Health Survey, Longitudinal Survey of Children and Youth, provincial health surveys); disease-based registries (e.g. cancer, diabetes, heart health); other registries (e.g. Registered Persons Data Base, Vital Statistics, workplace compensation/injury, registry of Status Indians); and non-health databases (e.g. employment, education, justice, environmental pollution/climate, fisheries, forestry). Users accessed encrypted individual-level data as well as aggregate-level data. Almost all users interviewed worked with linked data sets. All users interviewed for this study were senior established researchers/scientists; their views do not necessarily reflect those of junior researchers/scientists.

Data collectors/custodians (CCs) included government departments (5), regional/district health authorities (2), research data centres (7), disease registries (3), and other special purpose institutions (8). Five national organizations were represented within the total of 25 collectors/custodians, but due to small cell sizes, their responses cannot be reported separately.

2.3.1 Availability

Availability reflects the relationship between users' need for, knowledge of, ease of access to and ability to link data, and collectors/custodians' motivations, ability or authority to share data.

Users

Although a few users mentioned lacking knowledge of the existence of data sets, the more frequent comment concerned the unavailability of databases required for their investigations. Many spoke of the lack of standardized information on home care or long-term facility care at the individual level, and the lack of robust information on health human resources and physical measures, such as laboratory values on weight or blood pressure. Although there was said to be good information on the incidence of cancer and communicable diseases, there was a need for better data on cancer staging, chronic diseases, and dementia, to name a few.

Of existing data, a number of users indicated an inability to access data held by some provincial data centres or disease registries, or that access was restricted to those able to find scientists within the data centre/registry with whom to collaborate. Similar frustrations were expressed about access to databases held by some provincial governments that required research interests be relevant to government priorities. Others mentioned the proprietary attitude and unwillingness of researchers who collect their own data to share those resources with others.

While recognizing the importance of privacy legislation for the protection of the public and the individual, users referred to the “privacy chill” that hovers over the research community. Privacy restrictions were said to limit the level of data that can be obtained. For example, many users spoke of being able to get income data only at the neighbourhood level rather than at the individual level. Patient/client consent was also seen as a barrier to fulfilling research needs: e.g. the requirement to obtain patient consent to use their clinical data, to approach patients to request consent, or the inability to use data for purposes other than its original intent without consent.

Similar concerns were expressed about access to linked data sets, with the restrictions being more stringent.

Collectors/custodians

To appreciate the complexity of the above issues, it is necessary to understand the concerns of collectors/custodians in making their data holdings available to users. CCs were asked to whom they generally made their data holdings available and to whom access was denied. The variation in response was not so much contingent on the type of CC (provincial government, regional/district health authority, provincial data centre, disease/special purpose body, or individual researcher) but on the regional culture of data sharing. Approaches ranged from an open “public utility model” of data sharing to a more guarded “fortress model.” The public utility vision generally held that access and use of data holdings was publicly beneficial; those with the fortress view safeguarded their data resources and restricted access to those who could either find collaborators within the organization (data centres, disease/special purpose registries) or whose research question was of value to the organization (provincial governments, data centres, disease/special purpose registries). However, in this latter category, the access restriction of some data centres was reported to be largely the result of their licensing agreements with host governments. In the main, CCs of all types denied access to commercial agencies on the principle that the data must be used for a public purpose, and to some extent, access was also denied to contentious advocacy groups. Aggregate tables were often made available to a wider set of users than encrypted individual-level data, which was largely restricted to researchers and scientists with an institutional affiliation.

The lack of data-sharing agreements between jurisdictions and other data holders was viewed as problematic by some CCs, particularly government agencies. For example, the sharing of surveillance

data, based on a need-to-know policy, hindered public health units. The ability to compare data with other jurisdictions across regions or even across hospital authorities was another cited limitation. Both national and provincial CCs that obtained data from provincial governments spoke of the latter’s view that data was power and an unwillingness to share that power. One respondent lamented that this resulted in “huge data holdings, but zero information.”

Observations and suggestions

Data users recommended the development of a number of databases, described in greater detail in Section 2.5. Research agencies should facilitate discussions between researchers/users and collectors/custodians to improve access to data. In addition, resources should be allocated to encouraging and creating greater linkages across databases and across jurisdictions.

2.3.2 Accessibility

Accessibility is the relationship between the location where databases are housed and can be accessed, and the location of users, focusing on the opportunities and challenges this creates for users.

Users

Statistics Canada has appreciably increased opportunities to access its data holdings through its Research Data Centres, and the majority of users recognized this. However, access was more onerous for users located at either a university or in a province (Saskatchewan, Newfoundland, and Prince Edward Island) without a Research Data Centre. In these cases, appointments, travel time and extra cost was required. For those provinces without Research Data Centres, lack of funds to pay for Statistics Canada staff was said to be an issue. Users, even those who had ready access to Research Data Centres, highlighted the mismatch between the research process and that of accessing data holdings, with Research Data Centres not recognizing the vagaries

and trial and error nature of analyses that require the repeated revisiting of data. It was reported that these difficulties were compounded if users wanted to link Statistics Canada data with that from another source, such as provincial administrative data. The dual access processes often did not parallel each other, and led to long waits. All the above issues were also identified as relevant in accessing data from some provincial data centres.

Collectors/custodians

Although Statistics Canada Research Data Centres are not available in every province, their creation has substantially improved access. The cost of running the Research Data Centres is split with the provinces, with the latter funding the personnel to staff the centre. Similarly, the creation of provincial data centres was an attempt to extend access to provincial data. However, the ease and extent of access to data held by provincial data centres, as already stated, is dependent on the broader culture. For example, regardless of the desire of some data centres to liberalize access, their licensing agreements with the province may prevent it. Nevertheless, the requirement of on-site use lamented by users earlier were reported by these same CCs as a defensive strategy to safeguard the privacy and security of the data, as well as to ensure that users understood the nature, capability and proper use of the data.

Observations and suggestions

Users suggested that every province should have a Statistics Canada Research Data Centre. They further recommended the creation of more provincial data centres within a province to minimize the amount of time required to travel to sites and to break the monopoly over knowledge said to exist in some provinces. Some custodians reported that they are currently exploring ways in which to extend data access through the liberalization of access policies, or the re-negotiation of licensing agreements.

2.3.3 Accommodation

Accommodation refers to the relationship between the manner in which databases and access to them are organized by collectors/custodians for users and the users' ability to accept and meet these requirements.

The request and approval processes and the conditions set by collectors/custodians for data acquisition vary considerably. Each collector/custodian has their own review process, often with variations depending on the nature and sensitivity of the data sought. A formal request often requires an application with clearly stated research questions, objectives and methods, the nature of data sought, and documentation of approval by an institutional ethics board. It is usually further required that the objectives of the research be for a public purpose—e.g. for teaching or improving health status or the health care system—and that the applicant be affiliated with an accredited institution. These latter conditions, plus the requirement for an ethics review, are the basis on which commercial agencies are usually excluded.

Some custodians require that the research/question be congruent with their organizational priorities. For other custodians, such as provincial data centres or disease registries, access requires joint collaboration and publication with scientists in their organization. If access to data is granted, the conditions usually include use of the data only for the question cited in the application; restriction on storage of the data; consent of clients/patients if applicable; destruction, storage for some period of time, or return of the data upon completion of the research; and review of draft publications. Purchase of data is usually on a cost-recovery basis.

Users

Almost universally, users found the process of accessing data onerous and time-consuming. A wait of one to three years for access approval was not uncommon. A number of researchers complained that by the time their request was approved and they

received the data, the period of their grant funding had expired. One participant spoke of “ministries starving you to death waiting for permission.”

The difficulty of the access approval process was compounded by privacy and consent requirements, which were seen as slowing the process down or making it almost impossible to obtain the data. Researchers in jurisdictions with more restrictive privacy legislation felt disadvantaged in competing for grants. These frustrations were exacerbated when requiring access to multiple databases or databases from several jurisdictions, each of which would have their own processes, timelines, and sometimes conflicting privacy rules and legislation. One researcher told of an attempt to link data from two different sources, both of which had their own privacy rules and regulations. The two data custodians in this instance could not agree who “owned” the linked data set and therefore, whose privacy rules/regulations should apply.

Because of the sensitivity of linked data, some collectors/custodians (provincial ministries, provincial data centres) insist on doing the linkage in-house. Furthermore, in some jurisdictions, CCs also insist on conducting the analyses. A number of researchers stated that they would prefer to conduct the linkages and particularly the analyses themselves.

Users argued that because CCs do not put a high priority on providing data to others, a lack of trained staff to process requests results in bottlenecks and backlogs. There were also complaints that data are delivered in a non-user friendly format. The lack of data documentation was said to limit the user’s ability to understand and analyse the data. The requirement by most data custodians that the data be used for precisely specified purposes was said to be difficult to comply with without access to the data documentation. The requirement that research objectives coincide with custodian priorities was

viewed as narrow and insular. The requirement to find partners within the custodian organizations for collaboration and joint publication was viewed as unreasonable and proprietary.

Collectors/custodians

Collectors/custodians confirmed the above complaints from users were familiar. In jurisdictions with a fortress model culture, the provision of data for research was not viewed as a core function. For these CCs, provision of data must add value in furthering one of their own priorities. This was also true of most disease-specific registries and some special purpose registries regardless of jurisdiction.

A general lack of funding was reported by CCs to result in fewer technical staff to process and support requests, accounting for the long time delays experienced by users in gaining access, in the lack of support available for users, and the paucity of data documentation. Collectors/custodians, especially government ministries, spoke of the increasing demand placed on them by data collection, maintenance, and the introduction of new technologies. The adoption of new technologies, while said to offer the potential of greater ease of data collection and access, was viewed as time-consuming and usurping other priorities in data management.

The insistence of some custodians to link the data, to analyze the data, or to require collaboration with their own scientists was said to emanate, as stated earlier, from a fear that outside users do not understand the structure and limitations of the data, and will pose questions, make interpretations and report findings not supported by the data.

Observations and suggestions

Some users called for a pan-Canadian vision, a uniform, standardized policy for data access and a move away from multiple ad hoc arrangements. Others suggested the creation of a single data repository within each province whose sole purpose would be

to ensure access to quality data from all government sectors. A single data archive, it was said, would eliminate the proprietary rights over data deemed to be granted to some provincial data centres, and would allow government ministries to redirect their focus to their mandates. Other suggestions included the development of a universal format in which data are exported and better data documentation to make it easier for users to understand and use data.

From the collector/custodian point of view, suggestions varied from doing nothing, because they are not in the business of providing data for research, to liberalizing licensing agreements (for centres in receipt of provincial government data). However, most suggested that funding be increased to specifically account for the provision of data as one of their core businesses.

2.3.4 Affordability

Affordability refers to the relationship between the cost of data purchase for users and the availability of funding to collectors/custodians for data maintenance and research support.

Users

Although researchers/users recognized that granting agencies, namely CIHR, have allowed data purchase as a budget item in grants, there was still considerable concern expressed about the cost of purchasing both linked and unlinked data. Particular reference was made to the high cost of CIHI data or special-purpose Statistics Canada data. One respondent spoke of having to pay \$75,000 for air purity data. The cost for provincial data was also said to vary by jurisdiction. A few argued that after purchasing data, limited funding remained for analysis. The requirement by some custodians to destroy the purchased data upon completion of the research was viewed as both wasteful and inefficient, particularly if future research called for the same data.

Collectors/custodians

Collectors/custodians indicated that the majority of their resources went to the maintenance and updating of their data sets, i.e. “feeding the beast,” which included the conversion of paper records to electronic files. Others lamented that funding was often available for the creation of databases or the purchase of software, but not for ongoing operation and maintenance. or the purchase of hardware. As a result these expenses had to be supported from access charges.

From the CC perspective, funding was rarely sufficient for the creation and particularly the cleaning and maintenance of databases, the recruitment of adequately trained staff or the training of staff, and the development or purchase of new technologies or data documentation. Moreover, the commitment of funding on a year-to-year basis made it difficult to plan and adopt new technologies.

Observations and suggestions

Users recommended that public data should be free and available on the Internet as it is in the US.⁸

Collectors/custodians recommended multi-year funding, and the recognition of the costs of data maintenance and cleaning in funding. Technologies such as the electronic health record were seen as innovations that would decrease the work of data technicians in updating and cleaning records.

2.3.5 Acceptability

Acceptability refers to an understanding and acceptance by users and collectors/custodians of their different roles and responsibilities.

Users

Users spoke of a mismatch in cultures between users/researchers and collectors/custodians. This

⁸As provided on sites such as the National Center for Health Statistics (http://www.cdc.gov/nchs/products/elec_prods/subject/nmcues.htm) and the National Information Center on Health Services Research and Health Care Technology (<http://www.nlm.nih.gov/nichsr/>).

mismatch was particularly stark between researchers and some provincial government data holders. Despite government calls for evidence-based decision making, ministries were accused of not appreciating the value and role of research. Privacy, confidentiality and security of data, not research for a public purpose, were seen as the primary concern for most collectors/custodians. The pressure from governments and granting agencies on researchers to do policy-relevant research was at odds with the construction of barriers through privacy legislation. Users believed that the overly cautious approach of government agencies and other data custodians to privacy and confidentiality matters arose from the fear of scandal and embarrassment. The public's concern about the unscrupulous use of their private information and a general lack of understanding of the research endeavour fuelled this fear.

Collectors/custodians

Collectors/custodians also spoke of a poor understanding of the benefits of research from their own funders. They indicated that there was a general lack of leadership in government, which translated into a low priority given to research and data development and maintenance, and the collection of large amounts of data without the technical capacity to use, analyze or manage it. Collectors/custodians also felt that there was a general language barrier between users/researchers and information technology developers, resulting in reduced communication.

Observations and suggestions

Rather than focusing lobbying efforts on politicians who are said to already understand the issues, users recommended an education program to address the public's concerns about privacy and security and to demonstrate the value of research in improving the overall health of the population and the functioning of the health care system. Some believe that it would assist public understanding if privacy legislation distinguished between researchers and other data users such as commercial organizations. Others

recommended that the CIHR could facilitate discussions with collectors/custodians with the view to negotiating greater flexibility in access to data while providing assurances about protection of privacy. Overall, both users and CCs reported a need to better communicate and to build relationships and trust between each other.

2.3.6 Adequacy

Adequacy refers to the relationship between the suitability and quality of data for users and the capability of collectors/custodians to maintain/enhance data quality and develop data documentation.

Users

One of the barriers to access and use frequently mentioned by users was inconsistent data quality, especially for provincial administrative data. However, concerns about quality varied across jurisdictions and in relation to different databases. Quality issues included inaccuracy in collection, resulting in questionable reliability, validity and completeness of the data sets. Incompleteness was also said to result from tardiness in updates and changes in policy, which resulted in data loss. For example, due to provincial cutbacks, some municipalities began charging the public for birth registrations, which resulted in under-registration. Another example was the loss of information resulting from the introduction of alternative payment methods for physicians. A lack of standardized definitions and data collection methods within and across databases, across jurisdictions, and over time, also contributed to concerns about reliability and validity. Finally, many cited the difficulty of using data for purposes other than those for which they were originally collected.

Collectors/custodians

Collectors/custodians were very aware of the issues of data quality. Some reported spending a great deal

of time and resources on this issue; others spent only as much time and resources as funding would allow. For some, quality was akin to “fence-mending,” an ongoing struggle. For others, quality was less of a concern than developing and updating data documentation.

The following list of deficiencies reflects a composite of views from collectors/custodians and does not refer to all data. Frequently mentioned quality issues for administrative data resulted from inaccuracies introduced through errors in coding, provider gaming, or from converting paper records to electronic files. Inconsistent standards in data definitions and collection methods were another source of error. Respondents indicated that quality, especially in clinically-oriented data, was declining in some data sets over time. Other databases were incomplete or out-of-date, especially community data, outpatient data, or registered persons databases. Policies such as the voluntary submission of data, and changes in policy, such as the introduction of alternative payments for physicians, resulted in data loss and incompleteness. Additions or changes in variables—e.g. switching from ICD9 to ICD10, or changing regional boundaries—were not only time-consuming, but also a source of data error, making temporal comparisons difficult.

Quality issues for survey data included low response rates (and the consequent inherent response bias that introduces), missing data, and inconsistent data collection and reporting. Registry data were often cited as being incomplete or subject to provider gaming.

Each new generation of technology was said to create a risk of data loss. For example, the move away from print surveys to directly keyed-in responses in CATI systems, where data are captured in a system-dependent file, introduces potential loss unless these data are moved into migratable formats. Differing technology across jurisdictions, and the

lack of convergence in information technology, makes it difficult for CCs to collect information uniformly. The development and adoption of the electronic health record and the implementation of primary care reforms were viewed as huge steps forward in the improvement of coordinated, integrated and continuous care, as well as in data collection. However, concern was expressed that in the transition, data could be lost.

Major barriers to the improvement of data quality were said to range from a general lack of understanding of the importance of quality by the funders of data holdings; insufficient funding; the lack of data standards and data documentation; turnover of trained personnel and lack of trained personnel in the collection of data and the maintenance of databases; the number of “hands” that data must pass through; changes in technology and policy, which result in data loss; and consent requirements which introduce bias in participation.

Collectors/custodians employed many strategies to ensure data quality. Commonly used quality measures for administrative data included:

- data inspection for aberrant values;
- statistical testing, such as distributional consistency of data over time;
- logic checks;
- cross-validation with other data sources, chart audits/abstractions;
- user feedback; and
- training of data collectors.

Commonly used quality measures for survey or interview data included:

- methodological studies checking comparability of respondents and non-respondents;
- training and monitoring of data collectors; and
- the introduction of technology that automatically checks inconsistencies in responses.

Commonly used quality measures for registry data included:

- built-in edits and routines within the data collection process;
- triangulation with other databases; and
- retrospective reviews.

Observations and suggestions

Users and CCs recommended that data quality should be recognized as a priority and should be reflected in the resources apportioned to it. Standardization of data definitions and collection methods, increased training of personnel, education of data users, the development of better data documentation, and the introduction of technologies such as telephone-assisted surveys and electronic health records would all greatly improve data quality.

2.4 Findings—Electronic inventory

Both users and CCs were asked whether an electronic inventory of Canadian population-based health and health services databases would be a useful tool and what the characteristics and capabilities of such an inventory should be.

Forty per cent of all respondents said they would find an electronic inventory useful or very useful in helping them understand what databases and data sets were available, and reducing the time and effort spent searching for data. Others stated it would help jurisdictions advocate for the creation of local databases similar to those available in other jurisdictions, and would highlight the need for data standardization. A further 40% of respondents were equivocal and had reservations about the utility of an inventory, arguing that access and cost were the major barriers and not the knowledge of databases. For these respondents, an inventory would be useful if it also dealt with access, e.g. a one-stop shop approach. There was recognition that an inventory would probably be more helpful to junior researchers, but may be less pertinent to the community as a whole.

Others were concerned that an inventory would raise expectations and increase requests for access and therefore, the workload of custodians. A remaining 20% of respondents believed that an inventory would not be useful at all for the above reasons. For these respondents, supporting an inventory was beyond the scope of the existing resources of many custodians and money was better spent in capacity building, training and improving access.

The majority of both users and collectors/custodians provided either support or qualified support for the development of an inventory (88% and 75% respectively), though users were more likely to qualify their support (55% vs. 29%). In interpreting these results, we remind the reader that the sample of users interviewed were seasoned researchers, to whom knowledge of data sets was likely less problematic and for whom investments in resources may be seen as less important (See Table 2.2 for a comparison of responses).

In examining responses by region, respondents from western and eastern Canada were more positive about the utility of an electronic inventory (See Table 2.3).

If an inventory did exist, many respondents believed it should be user-friendly, web-based with a dedicated website, and searchable by key words and variables. Many felt it would be useful to have links to the data custodian's website, data dictionaries, data documentation, and links to articles/reports that used the data set. A number of respondents said it should be a portal for actual data access.

Most believe that the custodian of the inventory should be an independent, national body, either a new body with no other mandate created through a federal/provincial/territorial agreement or an existing national body, such as CIHI, the CIHR Institutes of Health Services and Policy Research and Population and Public Health, or Statistics Canada. A minority suggested an existing provincial university or research data centre.

Table 2.2: Usefulness of electronic inventory for collectors/custodians and users

Type of organization/ respondents	Yes	Qualified Yes	No	Total
Collectors/custodians				
Provincial government agency	3 (50%)	1 (16%)	2 (33%)	6 (99%) ²
National agency	1 (20%)	2 (40%)	2 (40%)	5 (100%)
Data centres/registries/ misc.	7 ¹ (54%)	4 (31%)	2 (15%)	13 ¹ (100%)
Collector/custodian subtotal	11 (46%)	7 (29%)	6 (25%)	24 (100%)
Users	6 (33%)	10 (55%)	2 (11%)	18 (99%) ²
TOTAL	17 (40%)	17 (40%)	8 (19%)	42 ¹ (99%) ²

¹Missing information for one respondent

Table 2.3: Usefulness of electronic inventory for provincial collectors/custodians and users by regions

	Yes	Qualified Yes	No	Total
West	8 (57%)	5 (36%)	1 (7%)	14 (100%)
Central	3 (21%)	7 (50%)	4 (29%)	14 ¹ (100%)
East	5 (56%)	3 (33%)	1 (11%)	9 (100%)
Total	16 (43%)	15 (41%)	6 (16%)	37 ² (100%)

¹Missing information for one respondent

²Due to rounding, total does not add to 100%.

2.5 Findings—Recommended investments

Users and CCs were asked to identify data gaps in population-based health and health services databases, and nominate where research-granting agencies need to place their strategic investments. Respondents had many suggestions for the creation of new data sets, including:

- Health services data—community care, mental health, public health, drugs
- Population health data—chronic diseases, disease staging data
- Biological and physical measurement data
- Longitudinal data—seniors, children, cohort studies
- Special populations—Aboriginal, homeless

Other suggested areas of investment included:

- negotiating standardized access and privacy policies;
- stable and ongoing funding for database maintenance and purchase;
- standardization of existing data sets and creation of data documentation;
- increased training for researchers to use large data sets and conduct secondary analysis, and for technicians to support large data sets;
- facilitating inter-regional comparisons and data linkage especially between health services data and determinants of health data;
- facilitating better understanding between collectors/custodians and users/researchers; and
- the development of a national vision and strategy for the collection, maintenance, and sharing of publicly funded Canadian data.

References

CIHI: Canadian Institute for Health Information. 2002. *Privacy and Confidentiality of Health Information at CIHI: Principles and Policies for the Protection of Personal Health Information and Policies for Institution-Identifiable Information*. Ottawa: CIHI.

CIHR: Canadian Institutes of Health Research. 2002. *Workshop on Privacy in Health Research: Sharing Perspectives and Paving the Way Forward*. Ottawa: CIHR.

CMA: Canadian Medical Association. 1998. Health Information Privacy Code. *CMAJ* 159(8):997–1006.

Ferris, LE. 2002. Industry-sponsored Pharmaceutical Trials and Research Ethics Boards: Are They Cloaked in Too Much Secrecy? *CMAJ* 166(10):1279–1280.

Kephart, G. 2002. *Barriers to Accessing and Analyzing Health Information in Canada*. Ottawa: CIHI.

Kluge, EH. 1996. The Role and Functions of Hospital-based Ethics Committees. *CMAJ* 154:1094–1095.

Tri-Councils: Medical Research Council (MRC), National Sciences and Engineering Council of Canada, Social Sciences and Humanities Research Council of Canada. 1998 (with 2000, 2002 updates). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Ottawa: The Councils.

Penchansky, R, JW Thomas. 1981. The Concept of Access: Definition and Relationship to Consumer Satisfaction. *Med Care* 29(2):127–140.

Tu, J et al. 2004. Impracticability of Informed Consent in the Registry of the Canadian Stroke Network. *New Eng J M* 350(14):1414–1421.

Willison D, et al. 2003. Patient Consent Preferences for Research Uses of Information in Electronic Medical Records: Interview and Survey Data. *BMJ* 326:373–377.

Willison, D. 1998. Health Services Research and Personal Health Information: Privacy Concerns, New Legislation and Beyond. *CMAJ* 159(11):1378–1380.

Chapter 3: Privacy and access issues in the use of population-based health and health services data

The trend in health privacy protection is towards an increasingly complex and constraining web of legislation. This creates logistical and compliance problems for researchers and others contributing to the development of data assets. On the other hand, future improvements in public health will increasingly depend on the more effective use of health data resources to monitor trends in health status, to investigate the causal roles of lifestyle, environmental and other risk factors within the degenerative diseases that increasingly account for morbidity and mortality, to measure and improve the quality and performance of health care services and to develop best practice for prevention and care.

(Magnusson 2002)

3.1 Introduction

Federal and provincial investments to create and support data custodians have increased over the last ten years in Canada and the intellectual and technical capacity to undertake research has grown exponentially. However, public concerns about the privacy of their information have also increased and many jurisdictions are creating new legislative and regulatory frameworks for the conduct of research using personal information (CIHR 2001).

These legislative and regulatory developments, which are also driven in part by international developments (notably in the European Union), take a number of forms. A number of provinces have passed specific health privacy legislation that applies to the public and private sectors, others rely on generic privacy legislation in the public and private sectors, and others depend on federal privacy laws.

Differences in how various Canadian regulatory frameworks deal with secondary use of health information for research purposes have led to

efforts to harmonize principles and standards across Canada. CIHR recently released draft guidelines for researchers for the protection of privacy and confidentiality of research data, with the aim of developing an approach that appropriately balances requirements for data access with the need to protect privacy (CIHR 2004). Research projects now cross international, national and provincial boundaries, and disciplines; health research is conducted in a variety of settings supported by a mix of public and private funds; and potential data sources for research are multiple and diverse. The increasingly cross-jurisdictional nature of health research makes it particularly important that harmonized or consistent regulatory frameworks be developed. Best practices or guidelines can be a useful tool for harmonization pending legislative harmonization.

A large body of literature articulates why the research product is a useful public interest enterprise (Black 1998; Black and Roos forthcoming; Detmer 2003; Emson 1994; Finkelstein 1999; Korn 2000; Lowrance 2003a; Magnusson 2002; Slaughter et al. 2003, 2004; Upshur et al. 2001). Some researchers have gone so far as to say that it is “morally reprehensible *not* to use available data to improve the health and wellbeing of the population” (Slaughter et al. 2004). In January 2004, Canada was a signatory to the OECD Declaration on Access to Research Data from Public Funding, which included a commitment to promote access to digital research data in the interest of advancing scientific innovation (OECD 2004).

Researchers do recognize, however, that using personal information to undertake research without the explicit consent of individuals needs to be discussed more openly than it has been in the past. They are urging the research community to look for new approaches that balance the risks and benefits to individuals and society as a whole (CIHR 2002b).

3.2 Organization of the literature review

This chapter reports on a literature search that was undertaken to supplement the views of the key informants presented in Chapter 2. The purpose of the literature review was to identify practical issues faced by researchers and data custodians around the use of Canadian population-based health and health services data for research. The theoretical arguments for protecting privacy or for the public support of health research are beyond the scope of this literature review.

A Medline search was undertaken to identify papers dealing with issues of access to data for secondary research, along with a survey of websites of data custodians and other stakeholders. Background materials prepared for two CIHR/CIHI-funded workshops on harmonizing research and privacy organized by the Institute for Clinical Evaluative Sciences and the Manitoba Centre for Health Policy were also reviewed. The literature review largely focussed on Canadian material. Details of the search strategy can be found in Appendix D.

The literature review identified a number of significant issues from the perspectives of data users and researchers, data custodians, legal and ethics scholars and privacy commissioners.

3.3 Issues identified in the literature

Eight primary issues were identified from the peer reviewed and grey literature. A discussion of each in relation to facilitating access to data while protecting individual privacy and ensuring confidentiality follows:

- Consent
- Data linkage
- Retention and destruction
- Security safeguards
- Review and oversight and the role of research ethics boards

- Multiple rules, policies and procedures
- Public communication
- Legal and policy frameworks surrounding data access and privacy protection

3.3.1 Consent

Many traditional confidentiality protections simply cannot cope with the complex data uses and flows in today's highly institutionalized, indeed industrialised, health care and research. Classic informed consent at each point of data handling for each purpose may be unduly onerous or impossible to obtain, and it may fail to inform legitimately and thus lack ethical validity.

(Lowrance 2003a)

According to legal scholars, the common law in Canada establishes a “very high standard for consent for the use of identifiable health care information” and that currently, “the legal obligation of informed consent in the research setting is tremendously strict and has been characterized as the most exacting duty possible” (Caulfield et al. 2003). Canada has developed what has been termed a “patient-centred” model: the patient remains in control of providing information with a clear right to withdraw consent. Secondary uses of data in this model require special permission and some mechanism is required to mediate competing interests (Magnusson 2002). To date, no Canadian cases have moved away from this model of decision making (Caulfield et al. 2003). But there is considerable debate over the circumstances when explicit consent is needed for the secondary use of personal information for research.

Though the single word “consent” is frequently used, there are at least three conceptual dimensions that form the various approaches to consent, including:

- *Opt-out vs. opt-in consent*—the former assumes that the information will be used unless the individual specifies that it not be used. The latter assumes that the information cannot be used unless the individual gives

permission to use it. A related dimension is that of *implicit vs. explicit consent*.

- *Broad vs. specific consent*—applies to explicit consent. Broad consent implies agreement to use of information for an array of research purposes. Specific consent refers to an individual's consent for use of only specified types of data and only for specified (types of) research and/or for research funded by specified types of organizations, etc.
- *Ability to withdraw consent at some future point in time*.

Generally, views in the literature can be grouped into three categories:

- those in favour of using information for research purposes without the explicit consent of individuals;
- those generally in favour of consent for usage, but recognizing that in some circumstances this may not be feasible or desirable;
- those in favour of providing information about usage and obtaining consent for use without exceptions.

All three perspectives assume that appropriate safeguards are in place to limit further derogations to individuals' privacy. Despite this, the rationales behind these perspectives are contestable and contested, leaving individual researchers confused about their ethical and legal obligations to obtain consent.

Secondary usage for research purposes without explicit consent

Arguments in favour of this level of consent include:

- The ability to use such data for the public good outweighs the claim of the right of the individual to give consent for each use, and if appropriate safeguards to protect privacy and confidentiality are in place, a socially sustainable system can be achieved (Bradburn

2001; Cayton and Denegi 2003; Emson 1994; Tu et al. 2004).

- The potential benefits to society of research greatly outweigh any hypothetical harm that access to personal information might entail (Finkelstein 1999).
- There is a difference between information and data and when information about an individual is separated from the identity of the individual, it should be thought of as data about "conceptual entities" (Bradburn 2001).
- Obtaining explicit consent for future use of information poses a number of challenges—it is not feasible to outline all the possibilities for use and obtaining blanket consent for an undefined future use is probably too general to have any legal authority (Caulfield et al. 2003; CIHR 2002b; Emson 1994).
- Using only the records of those who consent can lead to a biased sample and may skew research results and even cause harm if treatment decisions are made on this basis (Tu et al. 2004; Upshur et al. 2001).
- It is not clear that individuals actually want to give explicit consent every time their information is used (Upshur et al. 2001).

Secondary usage for research purposes with consent, and limited exceptions without consent

Arguments in favour of this level of consent include:

- If the information is specifically identified with individuals then they should have a right to determine how the information is used. But there will be cases in which a public need to use the information takes precedence over the desire of the individual for privacy (Bradburn 2001).
- Access to personal records should not require informed consent in certain circumstances; these circumstances need to be clearly outlined and exempted. The criterion of overriding public interest is proving to be too ambiguous (Peto et al. 2004).

- Consent should be required for the original research work but no further consent is required when data resulting from the study are used for other research projects, as long as the degree of confidentiality promised at the time of the original research is maintained (Bradburn 2001).

Secondary usage for research purposes only with consent

Arguments in favour of this level of consent include:

- Future use of research material can and should be permitted if there are strong systems in place to support privacy and confidentiality, and if a form of general and advance consent is developed and consistently applied (Cayton and Denegri 2003).
- Specific communities (for example, First Nations) have the right to own, control, access and possess research about themselves due to unique constitutional rights and distinct legal status (Schnarch 2004).
- Because informed consent is required for managing harm resulting from treatment, it is similarly seen as a necessary requirement to protect people’s privacy from unauthorized disclosure of information about them (Bradburn 2001).

A separate note is required for the issue of consent related to genetic material. Some legal scholars have argued that the existing consent norms “are incapable of accommodating much of the research associated with DNA data banks” and that adhering to well-established consent norms would prohibit a good deal of population research (Caulfield et al. 2003). These observers argue for a new model specifically for the use of genetic material for research purposes.

Caulfield et al. (2003) have proposed a data authorization model specifically for the collection of genetic information. This model uses a health

information directive to give participating individuals the ability to pre-specify uses, the ability to re-contact individuals, some rights of withdrawal, and rights of notification if commercial activity occurs. This model would require new legislative frameworks but is preferred to the current state of “fictional consent.”

3.3.2 Data linkage

Linking information from different sources at an individual level has long been accepted as a way to conduct population-based health research.

(Chamberlayne 1998)

Data linkage has been defined as the assembly of data in a common format from different sources but pertaining to the same unit of observation (United Nations 1999). Data linkage provides for much more powerful analysis than a single data set but raises a series of concerns beyond issues of consent. These include (Black and Roos forthcoming; GAO 2001):

- Heightened sensitivity whereby data sets that are relatively innocuous when separate are combined to create information that is more sensitive, e.g. by associating particular population groups or residential areas with certain health risks or diseases
- Special controls may be required when linking information that involves more than one agency, e.g. minimizing the amount of identifying information required to undertake the linkage or selecting a “trusted third-party” to do the linkage
- Where cross-agency linkage involves government agencies, additional steps should be taken to ensure that data collected for research are not used for administrative or program purposes (i.e. that the principle of functional separation is upheld)
- An increased risk that the combined information could identify actual individuals who were anonymized in the individual data sets.⁹

⁹A particular case cited a number of times involved a sample database containing only birth date and gender being linked to a voting registry. Twenty-nine per cent of the individuals could be identified when linked, and the number went up to sixty-nine per cent when general residential area was added in (Zoutman 2004).

A recent review of case studies of research projects using secondary data indicates there is a significant degree of variation in the access and linkage policies of data custodians and that the capacity and resources to undertake the work vary considerably (CIHR 2002b). The issue of data linkage requires further careful thought and documentation (Lowrance 2002).

3.3.3 Data retention and destruction

One approach to minimizing the chances of a privacy or confidentiality breach is to impose a requirement for destruction of data at the end of a research project. On the other hand, there is general agreement that data needs to be retained and archived for some period of time for audit purposes, and in some cases, for future use by researchers. However, there is no consistent approach to how the archiving should be done, the length of time the data should be retained, or protocols for future access (Lowrance 2002).

There is considerable debate about whether research data should ever be destroyed. Some argue that destruction would prevent expansion of research questions, reproduction of research results, and “look-backs” for public health reasons and that, in fact, it would be a waste of public funds to be forced to recreate what had already been publicly funded (CIHR 2002b). This approach is consistent with the OECD Declaration on Access to Research Data from Public Funding, to which Canada became a signatory in January 2004 (OECD 2004).

However, researchers cite the ongoing costs of maintenance, especially in ensuring continued privacy and confidentiality of retained data, as problematic. This relates to the general issue of the resource demands of maintaining privacy and confidentiality infrastructure. Resources for training, technical support, audits and security are often not built into grant applications or included in the calculations of administrative overhead (Slaughter et al. 2003). In response to these concerns, the

Canadian research community, including research funding agencies, is beginning to discuss how to better support data infrastructure and allow for compliance with emerging regulatory frameworks for the protection of privacy in research.

3.3.4 Security and safeguards

The issue of security and safeguards was a mandatory component of any discussion. Irrespective of the model of data custodianship, the approach to consent, or whether or not linkage was occurring, views about the need for tight safeguards were unanimous. Data custodians must be clear about what steps they take to safeguard the information they hold and must be transparent and accountable about their processes (CIHR 2004; Schnarch 2004). There are many options available, including using security clearances and employee confidentiality agreements, and employing physical and technical measures such as locked facilities, data encryption, passwords, access codes, tracking features and firewalls. Different groups use varying combinations of these features.

3.3.5 Review and oversight and the role of research ethics boards

It is clear that regardless of the data access model, there is consensus that some group must review research proposals to assess the trade-offs between the risks to individual privacy and the societal benefits of the research, and to ensure all possible steps are taken to maintain confidentiality. In some cases, this is a group or individual within the organization holding the data; in other cases, it may be a privacy commissioner external to the organization. Canada is increasingly looking towards research ethics boards to play this role in this area (Tri-Councils 1998; CIHR 2002a; CIHR 2002b; CIHR 2004).

While recognizing that research ethics boards are a logical mechanism for privacy reviews, a number of concerns have been raised. The boards themselves do not use standardized review criteria (Slaughter et al. 2003); are already under-resourced (CIHR

200b; Slaughter et al. 2003); need to strengthen their privacy expertise (CIHR 2002b); and are sometimes perceived to be too secretive (CIHR 2002b; Ferris 2002). Concern about how research ethics boards balance patient confidentiality against the benefits of particular types of research has been raised in other jurisdictions (Ward et al. 2004).

There is also a need to consider the existing mandates and powers of privacy commissioners and ombudsmen in overseeing privacy compliance by researchers under various regulatory frameworks. Any initiative to harmonize or make such frameworks more consistent must include the role of privacy commissioners and ombudsmen in order to promote effective privacy protection without inappropriately affecting research.

3.3.6 Multiple rules, policies and procedures

It is not one great beast that is arriving in Canadian physicians' offices but a menagerie of laws, regulations, guidelines, policy statements, consent forms and tool kits.

(Editorial 2003)

Perhaps the issue that causes the most frustration in the Canadian research community is the number of rules, policies and procedures that must be followed by those who don't have automatic access to data, those who wish to work with multiple data sets (and thus custodians), or those who wish to engage in cross-jurisdictional work.

In 2002, a research consortium wished to undertake linkage of the National Population Health Survey with health care utilization data in five provinces to determine whether lower income Canadians were less likely to receive preventative health services than higher income Canadians. They encountered so many barriers that they wrote a separate report specifically on the data access process (Kephart 2002). A number of issues were raised around privacy,

data quality and standardization, data extraction and linkage. The specific concerns related to privacy and confidentiality processes were:

- There were no standardized procedures, policies or criteria for privacy and confidentiality reviews in the provinces and with Statistics Canada;
- There was considerable variability in who actually conducted the reviews;
- The time required for the reviews varied;
- In addition to an ethics review and a privacy and confidentiality review, some provinces required other approvals for data exchange (in one province, sign-off from the Ministry of Health's legal branch was still pending at the time of report writing and the province was not included in the data analysis);
- Legal requirements of the custodian around what could be published and when sometimes ran counter to other contractual arrangements.

The same frustrations were articulated at a 2003 workshop of researchers and data custodians which was organized specifically to develop recommendations for a privacy "best practices" standard in Canada (Slaughter et al. 2004).

3.3.7 Public communication

Public understanding of where the data trail eventually leads and to what end and why must be improved if we are to expect them to engage in and understand the issues better than they are able to now.

(Cayton and Denegri 2003)

The importance of engaging the public in discussions about research, privacy and use of data was noted many times in the review material (CIHR 2002a; CIHR 2004; Cayton and Denegri 2003; Schnarch 2004; Slaughter et al. 2003, 2004; Upshur and Goel 2001; Upshur et al. 2001; Willison et al. 2003). A view held by many in the research community is that the

general public does not understand the importance of the research being undertaken and therefore needs to be convinced of its utility. As well, there is a view that being more transparent about the privacy safeguards in place will help to build public trust.

Detmer (2003) makes the point that much of public policy about privacy has been based on limited research from public opinion polls and surveys rather than studies of actual circumstances related to the collection of personal information. An innovative project is currently being undertaken by researchers from McMaster University and the Canadian Policy Research Networks to assess Canadians' attitudes towards the use of their personal information for health research. Using a variety of data collection methods, including a survey and eight dialogue sessions, this CIHR-funded project will probe public views on issues including the use, retention and disclosure of personal information and third-party data stewardship. This project represents a major advance not only in Canada, but internationally.

Another interesting research project currently underway in Canada is an attempt to develop better public communication tools (Upshur and Goel 2001). Researchers are developing a health care information directive that describes the various reasons for using personal information, what is done to protect privacy and confidentiality, and links the sensitivity of the data with the ethical appropriateness of its use. An evaluation of this tool found that consumers generally approved of the idea of the health care directive as a way of addressing the balance between the protection of privacy and the need for health research. But they questioned the feasibility of the directive in its current form, suggesting that it needed to be simplified and accompanied by public education. Many of the participants noted their appreciation for the inherent complexity of what the directive aims to achieve (Trace et al. 2004; Primary Care Research Unit 2003).

3.3.8 Legal and policy frameworks surrounding data access and privacy protection

Many in the research community may be inclined to assume that because they are doing research for the public good or because they have always collected personal information this way, they will be shielded from the privacy debate.

(Coy 2001)

While a full legal analysis is beyond the scope of this report, this section outlines recent legislative developments surrounding data access and privacy protection in Canada. Privacy is accepted as a principle of international law as part of the development of human rights and ethical standards to protect human dignity (CIHR 2001). Principles established by the OECD in 1980 have informed national statutes in many countries, including Canada. The *Personal Information Protection and Electronic Documents Act (PIPEDA)* is a federal private sector privacy law that regulates the collection, use and disclosure of personal information in the course of a “commercial activity”, including inter-provincial and international transfers of personal information. PIPEDA pertains only to commercial activities, and applies in any province that has not enacted substantially similar legislation for the private sector. This complex new environment has introduced ambiguities about the necessary steps needed for researchers to comply with privacy legislation, especially for cross-provincial work.

With some exceptions, PIPEDA requires an individual's consent before an organization can collect, use or disclose the individual's personal information for any purpose. It indicates that express consent is needed where “sensitive information” is involved; personal health information is defined in PIPEDA as sensitive information.

One exception to PIPEDA's consent requirement is for research — disclosure and use of personal

information for research is allowed without consent where the research purpose “cannot be achieved” without the identifiable information, it “is impracticable to obtain consent” and the organization first informs the federal privacy commissioner of the disclosure or use.

At the time of writing, four provinces have passed privacy legislation specific to health information.¹⁰ Other provinces have privacy laws that cover health privacy practices of public sector bodies and separate privacy laws that apply to the activities of private sector health organizations.¹¹

The existing regulatory framework in Canada appears to signal a very clear policy to permit secondary use of personal information for research purposes, but there is considerable variation in the practicalities of doing so. Like PIPEDA, provincial privacy laws address non-consensual disclosure and use of personal information for research purposes. Different provinces, however, set different conditions for when consent is not required, how approval (if any) for non-consensual use is obtained, and how information is to be used, held, protected and disposed of by researchers.

The former Medical Research Council (MRC), the Natural Sciences and Engineering Research Council (NSERC), and the Social Sciences and Humanities Research Council (SSHRC) developed a combined Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS) (Tri-Council 1998). Although the document does not have an enforcement mechanism in law, under a memorandum of understanding, institutions that receive grant funds from CIHR, SSHRC, or NSERC must adhere to the TCPS in all research conducted under that funding. The TCPS is therefore the de facto Canadian standard. Section three of the statement details the approach to privacy and confidentiality

as follows: “Without access to personal information, it would be difficult, if not impossible, to conduct important societal research which has led to major advances in knowledge and to an improved quality of life. The public interest thus may justify allowing researchers access to personal information, both to advance knowledge and to achieve social goals such as designing adequate public health programs.” Research ethics boards are suggested as a mechanism to judge consent issues and privacy trade-offs for specific studies (Tri-Council 1998).

As noted, existing regulations exhibit clear policy support for non-consensual use of personal information for research purposes, but there is a need to harmonize laws across Canada. The CIHR privacy best practices guidelines mentioned earlier are one step toward harmonization or consistency. It remains to be seen how these guidelines will work with or influence changes in the existing Tri-Council Policy Statement.

3.4 Emerging developments and suggestions for future work

From our review of the literature, we identified a number of potential actions for improving access to data for health research while strengthening privacy safeguards. Some of these activities are already underway in Canada. Five options are described further here and have informed the recommendations in Chapter 6. They are:

- Develop privacy tool kits;
- Develop best practice privacy guidelines or standards;
- Develop models of data stewardship;
- Strengthen and improve the practices of research ethics boards;
- Public communication about research and privacy trade-offs.

¹⁰Alberta, Saskatchewan, Manitoba and Ontario all have health privacy laws in force respecting public and private sector practices.

¹¹British Columbia and Quebec.

3.4.1 Develop privacy tool kits

The US General Accounting Office (GAO) and others have promoted the idea of a tool kit that research organizations and data custodians could use to protect privacy while allowing access to data (GAO 2001; National Committee for Vital and Health Statistics 2002). The onus is on the organization to establish best practices rather than a legislated directive to do so. A tool kit that could be employed by a number of data custodians would help to improve the standardization of practices.

A GAO report (2001) suggests the following tools for a privacy protection tool kit:

- Techniques for masked data sharing: list inflation, third-party models with three-way linkage procedure, group analysis instead of individual-level;
- Techniques for secure transfer of data: encryption, secure dedicated lines, separating identifiers;
- Written agreements and reviews of safeguards;
- Consent forms;
- Procedures to reduce re-identification: data “altering,” creating safe centres with control over access;
- Safer data: disclosure limitations, regrouping into categories, recoding some variables, licensing agreements;
- Security measures: access controls, audit trails of access, separate storage of data and keys to access data.

3.4.2 Best practice privacy guidelines or standards

Privacy approaches in Canada differ across jurisdictions and data custodians. As with other issues in Canadian health care, standardizing legislation across ten provinces, three territories and the federal

governments is a challenge. Where opportunities to harmonize legislation exist, they should be supported by the research community. Meanwhile, because it is the interpretation of legislation that shapes approaches to the use and access of data, the focus has moved to achieving harmonization via the development of best practices guidelines or voluntary standards for protecting privacy and confidentiality (CIHR 2002a; Slaughter et al. 2004). For example, a pan-Canadian Health Information Privacy and Confidentiality Framework has been developed by federal, provincial and territorial government officials with a view to creating a harmonized series of provisions to protect personal health information. These provisions would be consistent with PIPEDA and core provisions would apply in all jurisdictions. Particular provisions relating to the disclosure of personal health information for research purposes point to a central role for research ethics boards.¹²

Participants at the recent series of workshops on harmonizing approaches to dealing with privacy and research cited a need for guidelines or standards that are specific to health services and policy research. Existing guidelines (such as the Tri-Council Policy Statement) are based on health research in the form of clinical trials and are therefore an awkward fit with the types of data commonly used in health services and policy research (Weisbaum et al. forthcoming). The same workshops generated the following elements for a voluntary best practice standard for health services and policy research and proposed it be included as a separate section in the Tri-Council Policy Statement (Slaughter et al. 2004):

- Process descriptions: the purpose for use, required reviews and approvals for use including privacy commissioner or research ethics boards if appropriate, data transfer mechanisms, oversight and documentation, data protections, dissemination of an outcomes

¹²The framework was in draft form at the time of writing and was to be revised to reflect the results of extensive consultations.

- plan, appeals process if access is denied;
- Glossary: include local definitions matched to standard definitions from a statistical agency;
- Privacy code: include privacy impact assessment and confidentiality agreement;
- Consent: description of use of data without consent and why;
- Level of data: description of how researchers will use the least data possible with the highest level of aggregation;
- Roles: description of data custodians and privacy officers.

As a parallel development, CIHR's aforementioned draft guidelines for the protection of privacy and confidentiality in the design, conduct and evaluation of health research (CIHR 2004) propose the following:

- Determine the research objectives and justify the data needed: researchers should identify and document research objectives and include research questions which might become relevant after the initial data analyses, possible collaboration with other researchers or possible commercial uses;
- Limit the collection of any personal data: collect data at the lowest level of identifiability—minimizing as much as possible the collection or breadth of distribution of direct identifiers (name, street address) and other elements that could be used to identify an individual;
- Determine if consent from individuals is required: consent is fundamental for the use of personal data or human material when collected directly from an individual. When collected from other sources consent will generally be required unless the researcher can demonstrate why it should be waived. Under limited circumstances a waiver or partial waiver of a consent requirement may be permitted by law and approved by a research ethics board;

- Recruit participants in a manner that is non-invasive and does not lead to undue pressure on individuals to agree to participate;
- Inform prospective research participants about the research: provide full and frank disclosure of all information relevant to voluntary and informed consent and ensure that participants are given adequate opportunities to discuss their participation;
- Manage and document consent;
- Safeguard data confidentiality: data security measures should include organizational, physical and technological safeguards;
- Limit access to personal data: data sharing for research purposes is valuable but must be done with appropriate protections for privacy and confidentiality by controlling levels of access and having secure procedures for data linking, subject to agreements;
- Retain, destroy and archive data: data should be retained as long as is necessary to fulfill research purposes. Data may be destroyed or returned to the data provider if appropriate. Final research data sets should generally be archived for the use of the scientific community where resources exist, and with appropriate protections for privacy and confidentiality;
- Ensure accountability and transparency.

One of the key issues in developing these guidelines is how to determine when obtaining consent is not possible. Again, CIHR (2004) has proposed some criteria for this purpose :

- The size of the population group being researched is too large;
- For various reasons (time, lack of accurate information) it is too difficult to contact the individuals;
- There is a risk of introducing bias into the study if the actual consents are not randomly distributed;
- There is a risk of creating additional threats

to privacy if data needs to be reconnected to identify individuals to be contacted;

- There is a risk of harm associated with contacting the individuals;
- The requirement to obtain consent creates additional resources which “could impose a hardship on the researchers or organizations so burdensome that the research could not be done.”

3.4.3 Develop improved models of data stewardship

Some have suggested that the roles of data stewards are worth clarifying and developing further (Lowrance 2002; Slaughter et al. 2003). The idea of separating those holding the data from those undertaking the research with independent assessments of privacy risks and confidentiality protocols removes any actual or perceived conflict of interest (Flaherty 1992). Success would depend on a credible process, quick turnaround, and transparent decision making.

CIHR-funded work to identify and explore conceptual paradigms for the use of personal information in health research is currently underway and will support the development of improved data stewardship models by focusing on the application of consent concepts to the secondary use of health information.¹³

3.4.4 Strengthen and improve the practices of research ethics boards

Research ethics boards are one possible place in the Canadian environment to house responsibility for privacy reviews if the necessary supports are provided (Willison 2003). Further work is currently underway to identify innovative practices and delineate variation in policies and practices of REBs

in governing privacy, confidentiality and security issues in health research.¹⁴

3.4.5 Public communication about research and privacy

A number of suggestions have been made to facilitate public discussion about balancing research and privacy issues, including (Slaughter et al. 2003, 2004):

- policy documents and question and answer sheets that describe how data is collected and used;
- information on research organization websites informing the public about policies and procedures;
- standardized messages for the research community about effective and safe uses of personal information;
- templates for effective communication;
- relationship-building with reporters interested in health issues.

At the same time, some experts caution that the interests of researchers are not identical to those of the public (Willison 2003), since the research agenda encompasses commercialization, intellectual property and other interests. Thus, public communication efforts describing the benefits of research must take care not to conflate these interests.

Any investments to enhance access to and the use of health-related data in Canada must actively incorporate measures to protect privacy and confidentiality. The approaches described here can and are being used to support high-quality health and health services research while upholding the privacy values of Canadians.

¹³A description of this project by Gibson and colleagues can be found at www.cihr.ca.

¹⁴A description of this project by Willison and colleagues can be found at www.cihr.ca.

Bibliography

This bibliography includes cited references, and additional sources of background material used in the preparation of this chapter.

Amdur, RJ and MA Speers. 2003. A Practical Guideline for Identifying Research Intent with Projects That Collect Private, Identifiable Health Information. *Am J Clin Oncol* 26(3):e7–12.

Annas, GJ. 2002. Medical Privacy and Medical Research—Judging the New Federal Regulations. *N Engl J Med* 346:216–220.

Appelbaum, PS. 2000. Protecting Privacy While Facilitating Research. *Am J Psychiatry* 157(11): 1725–1726.

Bainbridge, WS. 2003. Privacy and Property on the Net: Research Questions. *Science* 302:1686–1687.

Baker, R et al. 2000. What Proportion of Patients Refuse Consent to Data Collection From Their Records for Research Purposes? *Br J Gen Pract* 50(457):655–656.

Banks, TM. 2000. Misusing Informed Consent: A Critique of Limitations on Research Subjects' Access to Genetic Research Results. *Sask L Rev* 63(2):539–580.

Birkett, N. 1993. Confidentiality and Research. *CMAJ* 148(4):486–487.

Black, C. 1998. Building a National Health Information Network. In *Canada Health Action: Building on the Legacy, papers commissioned by the National Forum on Health*. Sainte-Foy, Quebec: Editions MultiMondes.

Black, C and L Roos. (forthcoming). Linking and Combining Data to Develop Statistics for Understanding the Population's Health. Hunter E, Gibson R, Friedman D (eds) in *Health Statistics in the 21st Century: Implications for Health Policy and Practice*. Oxford, UK: Oxford University Press.

Black, N. 2003a. Privacy, Data and Health Research. *J Health Serv Res Policy* 8(Suppl 1):S1:1.

Black, N. 2003b. Secondary Use of Personal Data for Health and Health Services Research: Why Identifiable Data are Essential. *J Health Serv Res Policy* 8(Suppl 1):S1:36–40.

Boyd, P. 2003. Health Research and the Data Protection Act 1998. *J Health Serv Res Policy* 8(Suppl 1):S1:24–27.

Bradburn, NM. 2001. Medical Privacy and Research. *J Legal Stud* 30(2):687–701.

Buckovich, SA et al. 1999. Driving Toward Guiding Principles. A Goal for Privacy, Confidentiality and Security of Health Information. *J Am Med Inform Assoc* 6(2):122–133.

CIHI: Canadian Institute for Health Information. 2003. *Privacy Tool Kit*. Ottawa: CIHI.

CIHI: Canadian Institute for Health Information. 2002. *Privacy and Confidentiality of Health Information at CIHI: Principles and Policies for the Protection of Personal Health Information and Policies for Institution-Identifiable Information*. Ottawa: CIHI.

CIHR: Canadian Institutes of Health Research. 2004. *Guidelines for Protecting Privacy and Confidentiality in the Design, Conduct and Evaluation of Health Research. Best Practices. Consultation Draft*. Ottawa: CIHR.

CIHR: Canadian Institutes of Health Research. 2003. *Population-based Health and Health Service Data in Canada: Current Status and Future Health Research Potential*. Ottawa: CIHR.

CIHR: Canadian Institutes of Health Research. 2003. *Personal Health Information: Balancing Access and Privacy in Health Research. Summary, Recommendations and Follow up*. Ottawa: CIHR.

CIHR: Canadian Institutes of Health Research. 2002a. *Workshop on Privacy in Health Research: Sharing Perspectives and Paving the Way Forward*. Ottawa: CIHR.

CIHR: Canadian Institutes of Health Research. 2002b. *Secondary Use of Personal Information in Health Research: Case Studies*. Ottawa: CIHR.

CIHR: Canadian Institutes of Health Research. 2001. *Selected International Legal Norms on the Protection of Personal Information in Health Research*. Ottawa: CIHR.

Canadian Medical Association. 1998. Health Information Privacy Code. *CMAJ* 159(8):997–1006.

Caulfield, T et al. 2003. *White Paper on Genetic, Ethical, Legal and Social Issues in the Canadian Context*. Ottawa: CIHR.

Caulfield T, R Upshur and A Daar. 2003. DNA Databanks and Consent: A Suggested Policy Option Involving an Authorization Model. *BMC Med Ethics* 4(1):1.

Caulfield, T et al. Unknown. *The Canadian Lifelong Health Initiative and Legal Issues Regarding Privacy and Consent*.

Cayton, H and S Denegri. 2003. Is What's Mine My Own? *J Health Serv Res Policy* 8(S1):33–35.

Chamberlayne, R et al. 1998. Creating a Population-based Linked Health Database: A New Resource for Health Services Research. *Can J Public Health* 89(4):270–273.

Coy, KL. 2001. The Current Privacy Environment: Implications for Third-Party Research. *J Contin Educ Health Prof* 21(4):203–214.

Deerfield Group. 2002. *Identifying Health Infrastructure Priorities of Canadian Health Services, Clinical and Population Health Researchers: Beginning the Dialogue*. Workshop Report. Burlington, Ontario: The Deerfield Group

- Detmer, DE. 2003. Building the National Health Information Infrastructure Personal Health, Health Care Services, Public Health and Research. *BMC Med Inform Decis Mak* 3(10):1–20.
- Detmer, DE. 2000. Your Privacy or Your Health: Will Health Medical Privacy Legislation Stop Quality Health Care. *Int J Qual Health Care* 12:1–3.
- Donaldson, MS. 1994. *Regional Health Databases, Health Services Research and Confidentiality: Summary of an Invitational Workshop at the Institute of Medicine. National Implications of the Development of Regional Health Database Organizations*. Washington: Institute of Medicine.
- Editorial. 2003. Paying the PIPEDA. *CMAJ* 169:5.
- Emson, HE. 1994. Minimal Breaches of Confidentiality in Health Care Research: a Canadian Perspective. *J Med Ethics* 20(3):165–168.
- Federal/Provincial/Territorial Advisory Council on Health Infostructure. 2001. *Tactical Plan for a pan-Canadian Health Infostructure: 2001 Update*. Ottawa: Office of Health and the Information Highway, Health Canada.
- Ferris, LE. 2002. Industry-sponsored Pharmaceutical Trials and Research Ethics Boards: Are They Cloaked in Too Much Secrecy? *CMAJ* 166(10):1279–1280.
- Finkelstein, MM. 1999. Does the CMA's Privacy Code Go Too Far? Or Far Enough? *CMAJ* 160(6):781–784.
- Flaherty, DH. 1992. Privacy, Confidentiality and the Use of Canadian Health Information for Research and Statistics. *Can Public Admin* 35(1):75–93.
- Fraser, A. 2003. Privacy and the Secondary Use of Data in Health Research in Scotland. *J Health Serv Res Policy* 8(Suppl 1):S1:12–16.
- Gaze, B. 2003. Privacy and Research Involving Humans *J Law Med* 10(4):410–434.
- General Accounting Office. 2001. *Record Linkage and Privacy. Issues in Creating New Federal Research and Statistical Information*. Washington: US GAO.
- Government of Canada. 2002. *Privacy Impact Assessment Guidelines: A Framework to Manage Privacy Risks*. Ottawa: Government of Canada.
- Government of Canada. 2002. *Privacy Impact Assessment Policy*. Ottawa: Government of Canada.
- Hagey, J. 1997. Privacy and Confidentiality Practices for Research with Health Information in Canada. *J Law Med Ethics* 25:130–138.
- Higgins, J. 2003. The Patient Information Advisory Group and the Use of Patient-Identifiable Data. *J Health Serv Res Policy* 8(Suppl 1):S1:8–11.
- Hodge, JG et al. 1999. Legal Issues Concerning Electronic Health Information: Privacy, Quality and Liability. *JAMA* 282(15):1466–1471.
- Hoey, J. 1998. The CMA's Health Information Privacy Code: Does it go Too Far? *CMAJ* 159(8):953–954.
- Information Commissioner. Unknown. *Data Protection Principles*. London.
- Ingelfinger, JR and JM Drazen. 2004. Registry research and Medical Privacy. *N Engl J Med* 350(14):1452–1453.
- Kass, NE et al. 2003. The Use of Medical Records in Research: What Do Patients Want? *J Law Med Ethics* 31(3):429–433.
- Kazanjan, A. 1998. Understanding Women's Health Through Data Development and Data Linkage: Implications for Research and Policy. *CMAJ* 159(4):342–345.
- Kephart, G. 2002. *Barriers to Accessing and Analyzing Health Information in Canada*. Ottawa: CIHI.
- Kluge, EH. 1996. The Role and Functions of Hospital-based Ethics Committees. *CMAJ* 154:1094–1095.
- Korn, D. 2000. Medical Information, Privacy and the Conduct of Biomedical Research. *Acad Med* 75(10):963–968.
- Kosseim, P, ed. 2000. *A Compendium of Canadian Legislation Respecting the Protection of Personal Information in Health Research*. Ottawa: Canadian Institutes of Health Research.
- Kottow, MH. 1995. Confidentiality as Fair Agreement. *J Med Ethics* 21(2):117.
- Kulynch, J and D Korn. 2002. The Effect of the New Federal Medical Privacy Rule in Research. *N Engl J Med* 346:210–204.
- Lako, CJ. 1986. Privacy Protection and Population-based Health Research. *Soc Sci Med* 23(3):293–295.
- Little, M. 1999. Research, Ethics and Conflicts of Interest. *J Med Ethics* 25(3):259–262.
- Long, JA et al. 2001. *The CIHI Data Quality Framework, Version 1: A Meta-Evaluation and Future Directions*. Ottawa: Canadian Institute for Health Information.
- Lowrance, WW. 2003a. Learning from Experience: Privacy and the Secondary Use of Data in Health Research. *J Health Serv Res Policy* 8(Suppl 1):S1:2–7.
- Lowrance, WW. 2003b. Learning from Experience: Privacy and the Secondary Use of Data in Health Research. *J Biolaw Bus* 6(4):30–60.

- Lowrance, WW. 2002. *Learning From Experience. Privacy and the Secondary Use of Data in Health Research*. London: Nuffield Trust.
- Magnusson, RS. 2002. Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia's Health Information System. *Syd L Rev* 24(1):5–55.
- Manning, D. 2002. Commentary: Don't Waive Consent Lightly – Involve the Public. *BMJ* 324:1213.
- McCarthy, DB et al. 1999. Medical Records and Privacy: Empirical Effects of Legislation. *Health Serv Res* 34:417–25.
- National Committee for Vital and Health Statistics. 2002. *Shaping a Health Statistics Vision for the 21st Century*. Washington: National Centre for Health Statistics, Department of Health and Human Services.
- National Committee for Vital and Health Statistics. 1998. "Assuring a Health Dimension for the National Health Information Infrastructure." <http://ncvhs.hhs.gov/hii-nii.htm>.
- National Data Archive Consultation. Unknown. *Building Infrastructure for Access to and Preservation of Research Data in Canada*. Ottawa: Social Sciences and Humanities Research Council of Canada.
- National Data Archive Consultation. Unknown. *Phase One: Needs Assessment and Report*. Ottawa: Social Sciences and Humanities Research Council of Canada.
- National Task Force on Health Information. 1991. *Health Information for Canada*. Ottawa: National Health Information Council.
- Nelson, K et al. 2002. Do Patient Consent Procedures Affect Participation Rates in Health Services Research? *Med Care* 40(4):283–288.
- OECD: Organisation for Economic Co-operation and Development. 2004. *Ministerial Declaration on Access to Research Data from Public Funding*. Paris: OECD.
- Peto, J et al. 2004. Data Protection, Informed Consent and Research. *BMJ* 328:1029–1030.
- Primary Care Research Unit - Sunnybrook and Women's College Health Sciences Centre. 2003. *The Health Care Information Directive: Preliminary Results of Key Informant and Focus Group Evaluation of its Feasibility and Utility*. Ottawa: Health Canada.
- Ringwood, G. 1998. Toward Secure Health Information and the Electronic Health Record. *Health Law Can* 18(4):117–119.
- Robling, MR et al. 2004. Public Attitudes Towards the Use of Primary Care Patient Record Data in Medical Research Without Consent: a Qualitative Study. *J Med Ethics* 30(1):104–109.
- Roos, LL and JP Nicol. 1999. A Research Registry: Uses, Development and Accuracy. *J Clin Epidemiol* 52(1):39–47.
- Schnarch, B. 2004. *Ownership, Control, Access and Possession (OCAP) or Self Determination Applied to Research: A Critical Analysis of Contemporary First Nations Research and Some Options for First Nations Communities*. Ottawa: National Aboriginal Health Organization.
- Slaughter, P et al. 2004. *Harmonizing Research and Privacy: Standards for a Collaborative Future. Final Workshop Summary*. Toronto: Institute for Clinical Evaluative Sciences.
- Slaughter, P et al. 2003. *Harmonizing Research and Privacy: Standards for a Collaborative Future*. Toronto: Institute for Clinical Evaluative Sciences.
- Simon, GE et al. 2000. Large Medical Databases, Population-based Research and Patient Confidentiality. *Am J Psychiatry* 157:1731–1737.
- Sobel, R. 2002. No Privacy for All? Serious Failings in the HHS Medical Records Regulations. *J Biolaw Bus* 5(2):45–48.
- Squires, BP. 1992. Confidentiality and Research. *CMAJ* 147(9):1299.
- Tracey, CS et al. 2004. Feasibility of a Patient Decision Aid Regarding Disclosure of Personal Health Information: Qualitative Evaluation of the Health Care Information Directive. *BMC Med Inform Decis Mak* 4:13–20.
- Tri-Councils: Medical Research Council, National Sciences and Engineering Council of Canada, Social Sciences and Humanities Research Council of Canada. 1998 (with 2000, 2002 updates). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Ottawa: The Councils.
- Tu, J et al. 2004. Impracticability of Informed Consent in the Registry of the Canadian Stroke Network. *New Eng J Med* 350(14):1414–1421.
- United Nations Population Information Network. 1999. *Dictionary of Demographic and Reproductive Health Terminology*. New York: United Nations.
- Upshur, REG and V Goel. 2001. The Health Care Information Directive. *BMC Med Inform Decis Mak* 1(1):1–7.
- Upshur, REG et al. 2001. The Privacy Paradox: Laying Orwell's Ghost to Rest. *CMAJ* 165(3):307–309.
- Vandenbroucke, JP. 1998. Maintaining the Privacy and the Health of the Public. *BMJ* 316:1331–1332.
- Verity, C and A Nicoll. 2002. Consent, Confidentiality and the Threat to Public Health Surveillance. *BMJ* 324: 1210–1213.

Ward, HJT et al. 2004. Obstacles to Conducting Epidemiological Researcher in the UK General Population. *BMJ* 329:277–9.

Weisbaum, KM et al. Forthcoming. A Voluntary Privacy Standard for Health Services and Policy Research: Legal, Ethical and Social Policy Issues in the Canadian Context. *Alta Health L Rev*.

Willison, D. 2003. Privacy and the Secondary Use of Data for Health Research: Experience in Canada and Suggested Directions Forward. *J Health Serv Res Policy* 8 Suppl 1:S1:17–23.

Willison, D et al. 2003. Patient Consent Preferences for Research Uses of Information in Electronic Medical Records: Interview and Survey Data. *BMJ* 326: 373–377.

Willison, D. 1998. Health Services Research and Personal Health Information: Privacy Concerns, New Legislation and Beyond. *CMAJ* 159(11):1378–1380.

Woolf, SH et al. 2000. Selection Bias From Requiring Patients to Give Consent to Examine Data for Health Services Research. *Arch Fam Med* 9:nov/Dec:111–1118.

Zagon, S. 2001. *Asking Citizens What Matters for Privacy of Personal Information*. Ottawa: Canadian Policy Research Networks.

Zoutman, DE et al. 2004. The Confidentiality of Patient and Physician Information in Pharmacy Prescription Records. *CMAJ* 170(5):815–816.

Chapter 4: Existing inventory and data access activities

4.1 Introduction

One of the objectives of this project was to identify relevant conceptual frameworks that could serve as building blocks for an inventory of population health and health services research databases in Canada. When we started the project, it became obvious that there was a great deal of work going on, in Canada and internationally, in building such inventories. Most of this work is in the area of social sciences, and specifically in documenting and archiving survey instruments and survey data, but the models and frameworks were clearly relevant to this study.

The purpose of this chapter is to provide an introduction to some of the existing inventory activities. This lays the groundwork for the development of a proposed framework and data collection tool in Chapter 5, and also shapes an understanding of the need we have in Canada to better coordinate activities in this area—across disciplines and across organizations.

4.2 Methods

This part of the project involved an iterative process of discussions with key informants (including steering committee members), literature review (including web-based resources), and critical analysis and synthesis. Our review included some non-health inventories to understand the different nature of inventories across disciplines and to accumulate approaches for collecting information about the general content of databases and specific items or variables contained in each database in the inventory. Details on methods can be found in Appendix E.

4.3 The review

Our review of print and electronic literature,¹⁵ together with key informant interviews, identified a much greater than expected level of inventory activity underway in Canada and internationally, although none were specifically targeted at the combined areas of population health and health services research.

4.3.1 Data and documentation in the United States

In the United States, a number of agencies, including research funding agencies, support researcher access to data. In addition, accessibility of publicly funded research data is a basic policy principle of major US organizations. (See Appendix A for a review of the missions and mandates of selected organizations.) In January 2003 the National Science Foundation released a report from its Blue Ribbon Advisory Committee for Cyberinfrastructure that emphasized the risks of failing to deal with access issues in a more proactive way. Risks cited included data in different fields being left in irreconcilable formats, permanent loss of observational data due to lack of curated archives, and artificial barriers among disciplines resulting from uncoordinated development of incompatible tools and structures (Atkins et al. 2003). The Foundation has recently announced a Digital Archiving and Long-Term Preservation program with a goal to “stimulate research that builds capabilities for long-term management and preservation of digital materials”.¹⁶ This program responds to the recognition that as more research disciplines and social sectors rely on data-driven models and observational data for research purposes, there is an increasing challenge of archiving the relevant data. Current technologies are not up to the

¹⁵Much of the information provided below comes from websites of the various organizations involved in these activities. Quotations are used to identify where information is taken verbatim from these websites. In the absence of quotes, it should still be understood that this chapter often uses paraphrased material from the referenced websites. All cited websites were accurate as of December 2004.

¹⁶National Science Foundation, Digital Archiving and Long-term Preservation (DIGARCH), “Summary of Program Requirements,” <http://www.nsf.gov/pubs/2004/nsf04592/nsf04592.htm>.

task, so there is an imperative to create new methods of preserving historical material.

Within the National Institutes of Health, several institutes maintain inventories of databases for their respective research communities. As an example, the National Institute on Aging (NIA) maintains a database on longitudinal studies, which contains information such as contact information and web addresses for sources of longitudinal data on aging.¹⁷ The longitudinal studies, data sets and bio-specimen repositories documented in this inventory encompass a wide range of age groups (childhood to old age) and sources of longitudinal data. An online search feature enables researchers to select gender and age group of interest and to refine searches using multiple keywords. The primary purpose for establishing this database is to provide a resource for potential applicants for grants to the NIA. This is purely an inventory system; it does not provide access to aggregated data or microdata for research purposes.

The Agency for Healthcare Research and Quality (AHRQ), the primary US funding agency for health services research, supports data collection for a number of health services research databases as well as access to these resources by researchers. As an example, the Healthcare Cost and Utilization Project (HCUP) is a family of healthcare databases and related software tools and products sponsored by AHRQ and developed through a federal-state-industry partnership.¹⁸ These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of

treatments at the national, state, and local market levels. HCUP's objectives are to make these data available to a broad set of public and private users. Access is provided through an application process that includes a statement about proper protection and reporting of the data and payment of a fee that covers the "dissemination costs" for the data. The fees are minimal for the national sample databases (and nominal for students). State-level databases are available through a central application process, with fees that vary by state and the type (university, commercial, non-profit) of applicant.¹⁹

The National Center for Health Statistics, an agency of the US Centers for Disease Control, supports the collection of data and provides researchers with access to a number of data holdings, including the National Health and Nutrition Examination Survey, National Health Care Surveys, National Vital Statistics System, and the National Health Interview Survey through public-use access files and documentation available for download directly from their website.²⁰

The US National Library of Medicine maintains both a health services research and a public health information program to improve access to data and tools needed by researchers. It supports portals for topics in health services and public health, including information on current research projects, and data sets available for research purposes. An inventory that contains information on datasets and instruments in a standardized, searchable format provides contact information but no direct access to data.²¹

¹⁷National Institutes of Health, National Institute on Aging, Database of Longitudinal Studies, <http://www.nia.nih.gov/ResearchInformation/ScientificResources/LongitudinalStudies.htm>.

¹⁸Agency for Healthcare Research and Quality, Healthcare Cost and Utilization Project, <http://www.hcup-us.ahrq.gov/overview.jsp>.

¹⁹Fees for the state-level databases are significantly higher than for the national data, in some cases well over US\$1,000.

²⁰Centers for Disease Control and Prevention, National Center for Health Statistics, Data Warehouse, <http://www.cdc.gov/nchs/dataawh/ftpserv/ftpdata/ftpdata.htm>.

²¹United States National Library of Medicine, Health Services Research & Public Health Information Programs, <http://www.nlm.nih.gov/hsrph.html>.

4.3.2 Data and documentation in the United Kingdom

The United Kingdom has recently launched a number of high profile initiatives intended to promote access to data for research. The Economic and Social Research Council (ESRC), the UK's leading research funding and training agency in the area of economic and social research, has a "datasets policy" that must be adhered to as a condition of funding. This policy requires all applicants whose research proposal involves funding for primary data collection or for access to existing data sets, to establish in their application that the required data are not already publicly available. It also requires all award holders to offer for deposit copies of data collected as part of funded research, whether the data arise as a result of primary data collection, or are derived data sets resulting from ESRC-funded work. The data set must be deposited "to a standard which would enable the data to be used by a third party, including the provision of adequate documentation."²² ESRC allows applicants to include the time and financial costs of this work in their application for funding.

ESRC also supports the Economic and Social Data Service (ESDS), a national data archiving and dissemination service initiated in 2003 to offer enhanced support for the secondary use of data across the research, learning and teaching communities.²³ This service has responsibility for cataloguing and archiving, and provides access and support for an extensive range of economic and social data. The ESDS has been established as a distributed service, based on collaboration between four key university-based centres of expertise. These centres work together to provide preservation, dissemination, user support and training for an extensive range of

key economic and social data, both quantitative and qualitative, spanning many disciplines and themes. The overall direction and management for the ESDS is hosted by one of these four affiliated centres, the UK Data Archive (UKDA) at Essex. Deliverables for management of the ESDS include supporting a "universal portal" to provide an entry point to all ESRC-funded data services; links to other social science data services and resources, and multi-level searching across distributed data collections; general user support through establishment of a central "first stop" help service desk; standards and systems management through provision of procedures for data preparation, preservation, processing, and documentation, cataloguing and conditions of use; implementation of an acquisitions policy to ensure a coherent policy and collection; development of promotional materials and outreach activities; and interaction with an advisory committee of stakeholders to help strategic planning.²⁴

The UK Data Archive houses the largest collection of digital data in the social sciences and humanities in the UK and provides an electronic search catalogue and support for secondary use of quantitative and qualitative data in research.²⁵ It is responsible for acquiring, storing and disseminating machine-readable data (both quantitative and qualitative) generated as a result of ESRC funding. It collects information in a standardized manner for a large number of data resources, including information about conditions of access, availability and contact information. Data from the UK Archive, and from other partners in the Economic and Social Data Service, are online but are not available to the public. The service offered by this site is most similar to the National Center for Health Statistics in the United

²²Economic & Social Research Council, "Datasets Policy," <http://www.esrc.ac.uk/esrccontent/ResearchFunding/sec17.asp>.

²³Economic and Social Data Service, "Home," <http://www.esds.ac.uk/>.

²⁴Economic and Social Data Service, About ESDS – ESDS Management, "Detailed overview of service," <http://www.esds.ac.uk/about/manoverview.asp>.

²⁵UK Data Archive, "Home," <http://www.data-archive.ac.uk/home/index.asp>.

States, except that access to data is controlled and limited to researchers affiliated with universities and other “further education” entities in the UK. Users must register through their university, and with the username and password provided, have free access to data.²⁶ There is some information collected about the purpose for data acquisition, and a time limit for data use is applied, but this appears to be a requirement for information purposes for the data providers rather than a process of application and review.

The Medical Research Council (MRC) is the UK’s primary agency for funding health research. In 2001, the MRC established a data archiving and access project, with the objectives of consulting, reviewing case studies and developing recommendations concerning data archiving and access. In response, the MRC decided to promote collaborative data sharing and preservation in population research “by encouraging a life-cycle approach to data management, and by facilitating identification and development of generic informatics standards, tools and guidance.”²⁷ In addition, the MRC developed a draft “data sharing and preservation policy” with associated strategic goals and an implementation plan that outlines activities spanning the years 2003 to 2005. The draft policy, developed in 2002, states that the principle of data sharing in research is “the norm,” while recognising the special responsibilities

that researchers have in relation to people, their data and their legal and ethical rights.”²⁸ It suggests that building data sharing processes systematically into routine data management is good research practice because it strengthens quality, enables replication and audit, and limits duplication of efforts in the collection of data and construction of data sets.²⁹ The policy lays out an expectation that investigators supported by MRC funding will make explicit provision for sharing and preservation of data in the planning and execution of their research by making their research data available in a timely and responsible manner to the scientific community for subsequent research with as few restrictions as possible; explaining the distinctive, added value that will be provided in establishment of new or significantly extended large data collections; and that investigators requesting funds for new data collection will be required to include a data sharing and preservation plan in their proposals.³⁰

A UK initiative that provides useful understanding about data inventories to support researcher access is the Directory of Clinical Databases (DoCDat). DoCDat provides an inventory of (mainly) clinical databases existing in the UK, and for each database reviewed provides a brief description, a simple assessment of the quality of the data, and contact details for the custodian.³¹ DoCDat was started in 1999 by a group of researchers, clinicians and

²⁶Nominal charges apply where data are requested on a CD rather than through download from the Internet.

²⁷Medical Research Council, Science Strategy – Strategy Implementation – Other initiatives, “Data Sharing and Preservation,” http://www.mrc.ac.uk/index/strategy-strategy/strategy-science_strategy/strategy-strategy_implementation/strategy-other_initiatives/strategy-data_sharing.htm.

²⁸Medical Research Council, Science Strategy – Strategy Implementation – Other initiatives, “Data Sharing and Preservation,” http://www.mrc.ac.uk/index/strategy-strategy/strategy-science_strategy/strategy-strategy_implementation/strategy-other_initiatives/strategy-data_sharing.htm.

²⁹Medical Research Council, Science Strategy – Strategy Implementation – Other initiatives – Data Sharing and Preservation, “Draft MRC statement on Data Sharing and Preservation policy,” http://www.mrc.ac.uk/index/strategy-strategy/strategy-science_strategy/strategy-strategy_implementation/strategy-other_initiatives/strategy-data_sharing/strategy-data_sharing_policy-link.

³⁰This policy appears to have been heavily influenced by a draft report on data archiving and sharing written by individuals affiliated with the UK Data Archive. See Medical Research Council, “MRC Population Data Archiving and Accessing Project – Draft 2.0,” September 2002, http://www.mrc.ac.uk/index/strategy-strategy/strategy-science_strategy/strategy-strategy_implementation/strategy-other_initiatives/pdf-ukda_draft_report.pdf-link.

³¹Directory of Clinical Databases (DoCDat), “Home,” <http://www.docdat.org/>. Similar to many other organizations, DoCDat does not provide direct access to databases, but instead gives researchers contact information for data stewards or owners.

information specialists who developed a prototype structured questionnaire in some ways similar to the data collection tool developed for this project (Chapter 5). With the financial support of the Nuffield Trust, the tool was tested empirically in a pilot study of 24 UK databases. This resulted in modification of the questionnaire to improve the face and content validity of questions about database quality, and to reduce any floor or ceiling effects in the scoring system. The DoCDat website was launched in 2001 but its value was formally recognized by the Department of Health in its response to the Bristol Inquiry, which led to a commitment to the establishment of a directory of clinical audit databases in 2002. This provided funding for an expansion of the DoCDat assessment questionnaire, an expansion of the inclusion criteria for databases referenced by the resource, as well as enhancements such as improving search capability, enhancing indicators of data quality, and providing more comprehensive information about the resource. Ongoing maintenance and development of the resource are supported by the National Centre for Health Outcomes Development.

The Data Documentation Initiative

In the course of our work and interviews, we became aware of formalized approaches that have been developed by the international social sciences research community to provide metadata (data about data) about social science data research resources. While much of this work is related to a non-health content domain and to a narrow set of data resources (i.e. survey and statistical data), the purpose of this work is consistent with the aims of this project. The most advanced work in this area has been done by a project called the Data Documentation Initiative, which

started in 1995 to “produce a stable metadata standard for describing, finding, and using the survey datasets that underlie much of social science research.”³² The Data Documentation Initiative is an international collaboration interested in harnessing the power of the Internet to increase the sharing of research data sets, documentation and output. Its backbone is the emerging technical capabilities of the Internet, such as XML, or extensible mark-up language, that allow data to be captured, searched and displayed with an extraordinary amount of flexibility.³³ Ultimately, the Data Documentation Initiative is a consortium that will develop a set of standards to be implemented, in full or in part, on a voluntary basis, by interested parties across the world.

The work of the Data Documentation Initiative is about form rather than content. Its aim is to promote consistent cataloguing of information about research projects and data sets, even when such inventories might be hosted on many different websites across many countries. Its technical structure has been adopted by many of the leading inventory and data archiving efforts, including the Economic and Social Data Service and the UK Data Archive.

4.3.3 Data and documentation in Canada

Our review of Canadian literature, web-based resources and interviews revealed considerable activity in this area, with a number of Canadian agencies taking on active roles in enhancing access to data for researchers.

Statistics Canada identifies itself as is the “official source for Canadian social and economic statistics and products.”³⁴ Notably, however, it did not begin

³²Data Documentation Initiative, About the DDI Alliance, “Strategic Plan 2004-2006,” <http://www.icpsr.umich.edu/DDI/org/strategic-plan.pdf>.

³³The Data Documentation Initiative has been developed to be compatible with, or in compliance with, various international technical (such as ISO-11179) and bibliographic (such as Dublin Core) standards. There is a Canadian group of data librarians working to pursue further applications of DDI. The group is CANDDI, and the contact person is Chuck Humphrey at the University of Alberta.

³⁴Statistics Canada, “First visit to our site?,” <http://www.statcan.ca/english/about/first.htm>.

to focus on data collection directly related to health until 1994, with the implementation of the National Population Health Survey (NPHS).³⁵ Over the past two decades, Statistics Canada has attempted to respond to researcher needs with the development of a number of initiatives. The first of these is the Data Liberation Initiative, where participating academic institutions pay an annual subscription fee that allows their faculty and students unlimited access to numerous Statistics Canada public use microdata files, databases and geographic files.³⁶

Another program, the Research Data Centres (RDCs) program, is part of an initiative by Statistics Canada, the Social Sciences and Humanities Research Council (SSHRC) and university consortia to provide researchers with access, in a secure university setting, to microdata from population and household surveys.³⁷ The centres are staffed by Statistics Canada employees and are operated under the provisions of the *Statistics Act*. They are accessible only to researchers with approved projects who have been sworn in under the *Statistics Act* as “deemed employees.” The RDCs program has also recently introduced the Research Data Centre: Information and Technical Bulletin, a new forum for current and prospective RDC users to exchange practical information and techniques for analyzing data sets available at the centres. The bulletin is released in the spring and fall of each year.³⁸

As an alternative to the Research Data Centres, the Remote Data Access Program enables researchers to

write and test their own computer programs using a file with artificial data; they can send refined programs via the Internet to Statistics Canada, where they are run on the relevant microdata file.³⁹ In addition to these resources, Statistics Canada maintains information about definitions, data sources and methods on its website,⁴⁰ though it is not clear how actively the research community uses this resource.

These initiatives have led to a sharp increase in the use of Statistics Canada’s considerable archive of health-related data resources and data by researchers and have also been key in improving the cooperation between the creators of official statistics and the research community.

The Canadian Institute for Health Information (CIHI), with its data holdings in the areas of health services, health expenditures and health human resource, has recently developed specific processes for making microdata available to researchers, with a special program targeted to graduate students. It has developed materials that provide information for researchers and has made presentations to the research community. Its website includes a “request data” section targeted specifically at the research community.⁴¹ This resource provides information about CIHI’s data holdings (using a standard format), information about how to make a data request, an online data request form, and CIHI’s principles and policies for the protection of health information, which govern all data requests. There is no direct access to data from CIHI’s website. Requests for

³⁵The NPHS was launched as a longitudinal survey, meaning that the production of public use microdata files required stricter suppression of data elements to protect confidentiality than earlier cross-sectional resources (Wolfson 2004). Wolfson indicated that researchers therefore found the public use files developed from the longitudinal form of this resource to be very frustrating: the suppression rules meant that much of the rich context for analysis was missing.

³⁶Statistics Canada, “Data Liberation Initiative (DLI),” <http://www.statcan.ca/english/Dli/dli.htm>.

³⁷Statistics Canada, “The Research Data Centres Program,” <http://www.statcan.ca/english/rdc/>.

³⁸Statistics Canada, “The Research Data Centres information and technical bulletin,” <http://www.statcan.ca:8096/bsolc/english/bsolc?catno=12-002-XWE>.

³⁹Statistics Canada, “Remote Data Access: On-Line Service for Research and Analysis,” <http://www.statcan.ca/english/edu/rda/>.

⁴⁰Statistics Canada, “Definitions, data sources and methods,” <http://www.statcan.ca/english/concepts/index.htm>.

⁴¹Canadian Institute for Health Information, “Request Data,” http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=reqdata_e.

data are reviewed and must meet privacy and confidentiality requirements.⁴² Charges for data access are based on a per record payment schedule, with differential pricing for public and private sector requests (a program for graduate students provides the data without charge).

While the data and statistical agencies have made significant progress in documenting the content of their own data holdings, Health Canada has developed a number of issue-specific data inventories, including in the areas of injury, cardiovascular disease, and air pollution as it relates to health. These inventories differ from those of CIHI and Statistics Canada by providing documentation about data resources from multiple data holders. Inventory activity appears to be taking place in at least two of Health Canada's branches, and within branches, in many diverse centres of activity.

The Public Health Agency of Canada (incorporating the activities of the former Population and Public Health Branch of Health Canada) houses a Centre for Surveillance Coordination, which has developed or partnered in the development of at least two major database inventories. The first, relating to injury databases, includes a searchable web-based inventory of approximately 94 Canadian data sources reflecting injuries for all age groups. It was developed using a common data collection tool that identifies such things as the purpose of the data source, the injury-related content of collected data, data collection methods and data availability.⁴³ The inventory was developed as a tool to support the optimal use of existing data sources and is being used as a key step in constructing

a national injury surveillance network. This resource has been evaluated and updated based on feedback from public health professionals, providers of data and researchers.

A second effort is an Inventory of Federal/Provincial/Territorial Environmental and Occupational Data Sources and Surveillance Activities,⁴⁴ produced by the Centre for Surveillance Coordination in partnership with the Healthy Environments and Consumer Safety Branch of Health Canada. This inventory uses a similar but not identical data collection tool as the injury surveillance project. The goal of this inventory is to make available a wide range of existing data sources relating to environmental and occupational health and to support their optimal use by health professionals, researchers and policy makers.

A second centre in the Public Health Agency, the Centre for Chronic Disease Prevention and Control, houses an Inventory of Canadian Cardiovascular Disease Databases.⁴⁵ This inventory is available on the Internet and collects a different set of information about the data sets it covers, and is presented in a different format, from the inventories produced by the Centre for Surveillance Coordination. In all three cases outlined, the intent is to provide introductory information and contact details for each data set, not access to the databases themselves.

The Information, Analysis and Connectivity branch of Health Canada conducts work in the development of data systems and standards for producing metadata. Health Canada is also home to the "Data and Information Sharing" client, or DAIS. This

⁴²Canadian Institute for Health Information, "Privacy and Confidentiality of Health Information at CIHI," http://secure.cihi.ca/cihiweb/en/downloads/privacy_policy_priv2002_e.pdf. This process is a bit different from that used by HCUP in the US, where the submission provides information for data owners, but is not reviewed per se.

⁴³Health Canada, "Inventory of Injury Data Sources and Surveillance Activities," <http://www2.itssti.hc-sc.gc.ca/clf/clfinventory.nsf/Home?OpenForm&lang=E>.

⁴⁴Health Canada, "Inventory of Federal/Provincial/Territorial Environmental and Occupational Data Sources and Surveillance Activities," <http://www2.itssti.hc-sc.gc.ca/clf/hecsinventory.nsf/Home?OpenForm&lang=E>.

⁴⁵Public Health Agency of Canada, Centre for Chronic Disease Prevention and Control, "Cardiovascular Disease," <http://www.phac-aspc.gc.ca/ccdpc-cpcmc/cvd-mcv/ccdb/index.html>.

system uses underlying technical structures of the Data Documentation Initiative in an effort to provide desktop access not only to data, but also to summary analyses of those data, documentation that supports them, and the ability to produce some basic tabulations of data. The development of DAIS is an ongoing effort, aimed at employees of Health Canada.⁴⁶

In summary, while Health Canada appears to have the greatest experience in developing inventories of databases across data custodians, to date many of the inventories have been independently developed by different centres within Health Canada. The inventories that have been developed typically use varying templates for collection of information about the databases. Moreover, it is not clear how well known or useful these inventories are to the research community and there is currently no single portal for the research community to gain access to them. To address the first issue, the Infostructure Development Division within the Centre for Surveillance Coordination is currently undertaking a project to review and compare content domains of a number of inventories (Sherman and Vaillancourt, 2004).

Administrative data sets collected by provincial ministries responsible for providing health and social services are also of great importance to population health and health services researchers. These departments vary in their approaches to making data and information about data available to researchers. Some interact with researchers directly to provide access to anonymized data for approved research projects. For example, the Régie de l'assurance

maladie of Quebec maintains a site that provides information about databases held, the contents of those data holdings, the process to review requests, and contact information for requesting access to data holdings.⁴⁷ Other provincial health ministries, such as Manitoba Health and the Ontario Ministry of Health and Long Term Care, have developed active working relationships with health service and clinical researchers who access the data within secure research settings.⁴⁸ In yet another model, the British Columbia Ministry of Health Services makes a selection of its data holdings available via the BC Linked Health Database, which in turn provides access to linked data for research projects that have received approval from data custodians.⁴⁹

In terms of research funding agency support for data sharing and archiving, the Social Sciences and Humanities Research Council of Canada is the only national funding agency in Canada with an explicit archiving and sharing policy. It operates on similar principles to the National Science Foundation and the National Institutes of Health in the US.⁵⁰ Canada has committed to work toward the establishment of access regimes for digital research data arising from publicly funded research projects as a signatory to the OECD Declaration on Access to Research Data from Public Funding.⁵¹ The National Research Council has been asked to lead a National Consultation on Access to Scientific Research Data (NCASRD) in conjunction with the Canada Foundation for Innovation, the Canadian Institutes of Health Research, Industry Canada, the Natural Sciences and Engineering Research Council, and the Social

⁴⁶See, for example, Bradley (2004) for more information.

⁴⁷Québec Government, Régie de l'assurance maladie du Québec, "Databanks," <http://www.ramq.gouv.qc.ca/en/statistiques/banques/index.shtml>.

⁴⁸See the Manitoba Centre for Health Policy (<http://www.umanitoba.ca/centres/mchp/>) and the Institute for Clinical Evaluation Sciences (<http://www.ices.on.ca/>).

⁴⁹Centre for Health Services and Policy Research, "About the BCLHD," <http://www.chspr.ubc.ca/Bclhd/aboutbclhd.htm>.

⁵⁰Government of Canada, National Consultation on Access to Scientific Research Data, "Report on the National Consultation Event Design Task Force Meeting," http://ncasrd-cnadr.scitech.gc.ca/tdfdesignreport_e.shtml.

⁵¹Organisation for Economic Co-operation and Development, "Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué," http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html.

Sciences and Humanities Research Council. The overall objective of the NCASRD is “to help Canada maximize the value received from its publicly funded natural and medical sciences research by recommending an appropriate framework, including guidelines, which will facilitate open and long term access to data coming from that research.”⁵²

4.4 Overall findings

Our review of international literature and web-based resources revealed that there is already a significant amount of work underway to enhance access to research data in the US and the UK. This work involves the development of data archives, inventories of databases, and portals aimed at researchers, as well as a number of complementary activities that support developments in this increasingly active area. Our review also suggested that, while there is no single best way to provide information about data, there are a broad variety of approaches and some evolving best practices. None of the best practice models were specifically targeted to the areas of population health and health services research, which cover a vast set of content issues and rely on a variety of data sources. Most efforts so far have concentrated on the documentation of survey data, often in combination with providing actual access to those data (for example, the UK Data Archive). In fact, the exploration we conducted suggests that there is a complex and evolving scientific agenda related specifically to the documentation of data and research resources. It is clear that to do the work of building an inventory well, it would be necessary to bring in additional scientific perspectives and resources.

In Canada, most of this activity is organization-specific and aimed at documenting organizational data holdings (e.g. Statistics Canada, which has focused primarily on documenting surveys; and CIHI, which has developed a format for providing information across a range of data holdings). Health

Canada has developed some topic-specific inventories that relate primarily to the area of population health, but there is little standardization across the approaches that have been developed and it is not clear how well known or how useful these products are to the research community.

While there is currently a great deal of activity, there is little coordination in the areas of improving data documentation and access. There is even less coordination of activities designed to serve the population health and health services research communities. There is no single portal to identify data sources and no standard format being used to compile information, the sources are in varying states of maintenance, and there is spotty coverage by agency and only narrow topic-specific information of variable quality (i.e. some of it is not up-to-date, some is of limited utility for the research community, and most is difficult to find). Finally, there is little sustained effort in the area of administrative data, an important (actual and potential) resource for population health and health services researchers. In short, there is no coordinated focused development that could provide a strong foundation for Canada’s research community.

References

- Atkins, DE et al. 2003. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Arlington, Virginia: National Science Foundation.
- Bradley, WJ. 2004. From Data to Knowledge: Standards and systems for sharing data across organizations. Paper presented at United Nations Statistical Commission and Economic Commission for Europe Conference of European Statisticians, Geneva, February 9-11, 2004.
- Sherman, Gregory and Alain Vaillancourt. 2004. Interview by Charlyn Black and Cathy Fooks. August 18.
- Wolfson, Michael. 2004. Interview by Charlyn Black and Cathy Fooks. August 27.

⁵²Government of Canada, National Consultation on Access to Scientific Research Data (NCASRD), “Home,” http://ncasrd-cnadr.scitech.gc.ca/home_e.shtml.

Chapter 5: Building an inventory of population health and health services research databases for Canada

5.1 Background

The funding agencies for this project identified a need to describe “the current status of population-based health and health services databases in Canada that are being used and show the potential for use in innovative and important health research” (CIHR 2002). The request for proposals (RFP) noted that while Canada has some of the best-developed data repositories for studying health and health care, “the challenge now lies in enhancing access to and use of current data infrastructure for the purposes of conducting important health research and to make wise investments to increase data and analytic capacity. However, investments to enhance data infrastructure for health research in Canada could be guided by a better understanding of current capacity and issues regarding access to and use of data across the country” (CIHR 2002).

Previous chapters in this report dealt with stakeholder perceptions about data availability, background issues of access, privacy protection and national and international activities around the development of inventories. This chapter concentrates on the development of tools for making information about relevant databases more widely available to the research community.⁵³

The first set of activities involved the development of a conceptual framework and taxonomy that could serve as a data collection instrument for the development of an electronic inventory of databases. The specific tasks for this component involved:

- 1) reviewing existing conceptual frameworks, taxonomies and projects that have been used to develop inventories of population-based health and health services research databases;

- 2) based on this review, development of a conceptual framework (or data collection tool) that provides useful information and allows the systematic categorization of data resources for funders, researchers and research users; and
- 3) testing of this data collection tool on a small sample of databases chosen from the fields of population health and health services research.

The second set of activities concentrated on identifying future options for creating a comprehensive electronic inventory of databases, using the data collection tool as the basic infrastructure. The specific tasks for this component involved:

- 1) identification of best practices in the development of electronic inventories;
- 2) outlining the principles for building electronic database infrastructure; and
- 3) identifying specific options for building an electronic inventory of Canadian population health and health services research databases, one that builds on the framework developed in Part I.

This chapter is broadly divided into three sections—methods, tools developed, and findings and lessons learned. The methods section provides an overview of the processes we used to conduct this piece of the project. The tools developed section presents two products developed for this project. These represent pilot versions of the major tools that would be required for development of an electronic inventory: the conceptual framework that ultimately would serve as a data collection tool for populating an electronic inventory of databases, and a framework to screen databases for inclusion in such an inventory.

⁵³The tasks for this portion of the project comprise activities defined in Parts I and III of the RFP and the funded proposal.

The findings and lessons learned section provides background information to understand best practices for development and maintenance of a Canadian electronic inventory. It also outlines lessons learned from the development of the conceptual framework as well as observations from pilot testing the data collection tool with a representative sample of databases. From this set of activities, we identify a set of potential models that could be used to build an electronic inventory of Canadian population health and health services research databases.

5.2 Methods

The methods for this part of the project involved a highly iterative process of literature review (including web-based resources), discussions with key informants (including steering committee members), and critical analysis and synthesis activities. Two tools were developed to support the creation of an electronic inventory; a data collection tool to collect information about databases, and a framework to screen databases for inclusion in an inventory. The data collection tool was pilot tested on five databases, providing the first entries for an electronic inventory. A detailed description of the methods is provided in Appendix E.

5.3 Tools developed

5.3.1 Conceptual framework and data collection tool

Developing a conceptual framework to identify content areas of data sets

A number of conceptual frameworks have been developed to guide research in the area of population and public health (i.e. CIHR's pillar four research) by providing a richer understanding of the many factors that influence the health status of individuals and populations. Perhaps most notable among these is the depiction of individual, system and social determinants of health developed by Evans and Stoddart (1990). A

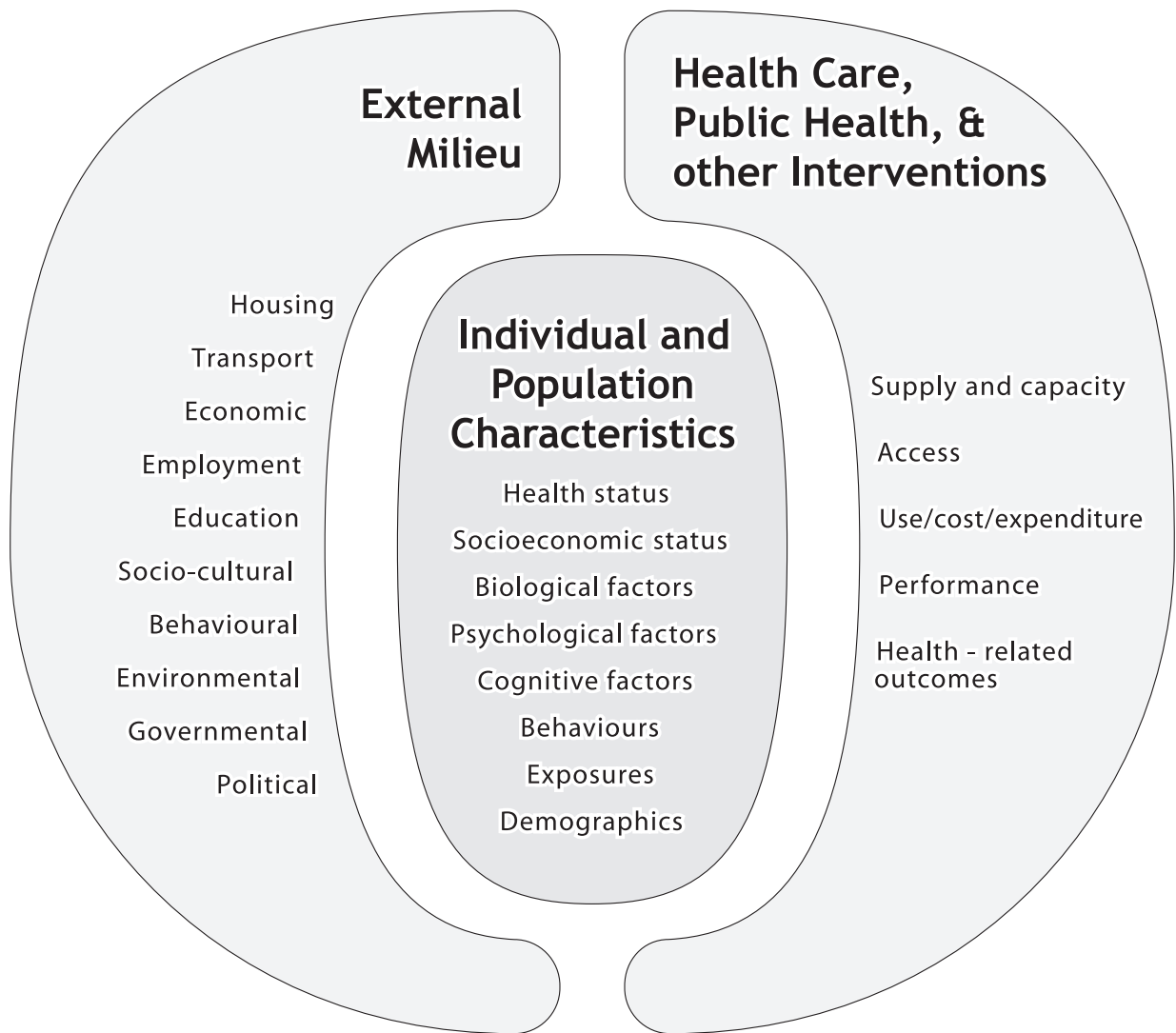
number of conceptual frameworks have been developed in the area of health services and policy research (i.e. CIHR's pillar three research), some less well known (Roos et al. 1995) and others that are well known but less comprehensive (Donabedian 1966, Aday et al. 1993). There is, however, no existing combined conceptual framework that adequately covers these two major areas of research activity (i.e. pillars 3 and 4 combined). Because of our need to map data holdings for their utility for work in both population health and health services research, we developed a conceptual framework that combined elements of both. This conceptual framework, in turn, served as a means of organizing a section of our data collection tool that focuses on mapping content of data holdings.

The resulting conceptual framework comprises three broad domains (Figure 5.1). Given the centrality of health states to both population health and health services research frameworks, we felt it was important to feature this aspect prominently, along with other characteristics of individuals and populations. Individual and population characteristics are featured as central in the resultant conceptual model. Included here are all the things that might describe an individual or population, such as age and sex (demographics), genetic predispositions (biological factors), early childhood experience, and educational attainment (socioeconomic status).

Consistent with population health frameworks (Evans and Stoddart 1990; Frank 1995⁵⁴), a second domain, depicted on the left side of the figure, identifies factors in the external milieu that are known to influence health (and to be related to other characteristics of individuals and populations). Included here are the characteristics of the surrounding community, either derived from aggregations of individual-level information, such as average educational attainment, or measured at the area level, such as the availability of recreational facilities and green space.

⁵⁴See also, Statistics Canada Health Template software (A framework for the collection and analysis of health information), <http://www.statcan.ca/english/spsd/helthtem.htm>.

Figure 5.1 Conceptual framework for documenting content domains of data holdings



To identify content domains for this area, we focused on two primary resources. The first was Statistics Canada’s Health Template software⁵⁵, which was developed as part of the National Task Force on Health Information. It provides an idealized comprehensive conceptual structure for health information as informed by a population health perspective. This framework involves interactions between a population of individuals, their external environment, and sets of

potential health-affecting interventions. The second major resource was a recent article that formally attempted to conceptualize community contextual influences on population health (Hillemeier et al. 2003). This study used a consultative process and literature review to identify 12 overarching dimensions of contextual characteristics that affect community and population health, as well as specific subcomponents related to each dimension.

⁵⁵Statistics Canada Health Template software (A framework for the collection and analysis of health information), <http://www.statcan.ca/english/spsd/helthtem.htm>.

The third domain, depicted separately in the area on the right side of the figure, relates to health-affecting interventions, including health care, public health and other deliberate interventions targeted at improving health. A number of frameworks were used to develop subcomponents of this area (see for example, Black 1998, Frank 1995). Included here are public health and health care interventions aimed at either individuals or populations, ranging from public health inspections (use/cost/expenditures), the availability of equipment (supply and capacity), patient satisfaction (performance), and health-system related outcomes.

These three domains were used to group sub-areas of content that were identified from further review of the literature. These domains and the sub-areas were then used to develop a format for data collection about the content of a given data holding. To ease collection of data, a tick-box format was developed so the data content could be mapped in a comprehensive and consistent manner across data holdings.⁵⁶

Developing other areas of the data collection tool

In addition to being able to capture information about the content area of data sources, an inventory must provide standard information about a given data set, such as how it is maintained, the form in which data are collected or available, and how researchers can access the resource. We relied on the review described in the previous chapter to help develop this area, with particular attention to innovative projects such as the UK Data Archive, the Data Documentation Initiative, and the Directory of Clinical Databases in the UK. There was little empirical work to underpin the development of these areas of the conceptual framework, but others have noted that some thoughtful common sense and consensus building are likely sufficient (Black 2004). We developed suggested questions for these areas by identifying and grouping

content areas from other initiatives, and synthesizing what appeared to be relevant approaches, modified as appropriate. In addition, we identified areas where new content was required because of the unique aspects of the population health and health services research domain.

The final data collection tool

By developing a conceptual framework that could consistently identify content areas of data sets, together with items that collected data on other aspects of data sets important to researchers, we developed a data collection tool that can be used to collect information about individual data holdings (i.e. metadata). The full data collection tool, which is quite lengthy, is included in Appendix F; an abbreviated outline of its major contents is provided in Figure 5.2. As outlined in this figure, the tool collects information about four major areas: general information, attributes of the data, data contents and data availability. The general information area captures details such as the purpose of the database, a general description of it, which organization is responsible for the data, what the reference population is and what time period the database covers. The attributes of the data capture more detail about the data set itself, including the data source (survey, census, administrative data), the unit of observation (individual or aggregated) and the potential for linkage to other data sources. The section on data availability provides information about whom researchers should contact for more information, and other details such as potential wait time for data and potential cost.

This tool would ultimately allow for the collection of data about data sets, the activity central to development and maintenance of an inventory. It represents a compilation of items adopted directly, or modified, from a variety of sources, together with content developed specifically by the research team.

⁵⁶An alternative would be to request delineation of content areas for a given data holding in an unstructured manner, a practice of many inventories. We felt it preferable to develop a format that would encourage consistent review of each data holding across relevant domains.

FIGURE 5.2: ABBREVIATED DATA COLLECTION TOOL FOR REVIEW OF DATABASES

★ Denotes mandatory field

1. GENERAL INFORMATION

- ★ Database Name:
- ★ General description:
- ★ Purpose:
- ★ Sponsor/Collector/Custodian:
Describe the timeframe covered:
General description of reference population and/or geographic area covered:
Data collection methods:
Changes in data over time/data updates:
Outputs (including analysis, reports and publications):
Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.):
Approximate number of records in database per year:
When and how often are data collected and how are data made available?
Language:
Funding agency and grant number (if applicable):

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

- ★ Data source:
- ★ Representativeness/population coverage:
- ★ Temporal nature of data:
- ★ Level of information collected/unit of observation:
- ★ Potential for data linkage (check highest level possible):
- ★ Years covered/available: _____

3. DATA CONTENT

- Individual/population characteristics:
- External milieu/factors that influence health
- Health care, public health and other health interventions

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

- ★ Contact:
- ★ Organization housing or maintaining the data source:
Associated link/URL:
Protocols that govern access to data:
Is there a data request process/form for researchers?
- ★ Is a data dictionary available?
Service charges:
Timeline to access data:
Training and support available for researchers:
Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers

5.3.2 Tool to identify scope of databases for inclusion in an inventory

One important aspect of developing an electronic inventory of existing data resources is the necessity of developing a set of criteria to delimit the scope of the resource. Specifically, there is a requirement for a mechanism to determine which data resources should be documented and which should not be listed in the data resource. This is particularly important for the current project, which is intended to develop a framework for both health services and population health research data holdings. The implication is that the framework is expected to function for administrative data on the use of health care services, registries of people with specific health conditions or those who are eligible for particular services, surveys of individuals identified through clinical or random samples, demographic information, educational attainment, use of a myriad other social services, and information on families, communities and neighbourhoods that will be collected and analyzed at an ecological (rather than individual) level.

Even though the scope is broad, there are clear and consistent boundaries that can be drawn. Using an inventory of data sets that was created in Alberta as a starting point (Alibhai, McCaffrey, and Saunders 2002), we developed the framework presented in Figure 5.3 to define these boundaries. This framework serves as a decision tool for selection of relevant databases for inclusion in the proposed inventory.

5.4 Findings, lessons learned and potential models

5.4.1 Lessons learned from developing and testing a prototype data collection tool

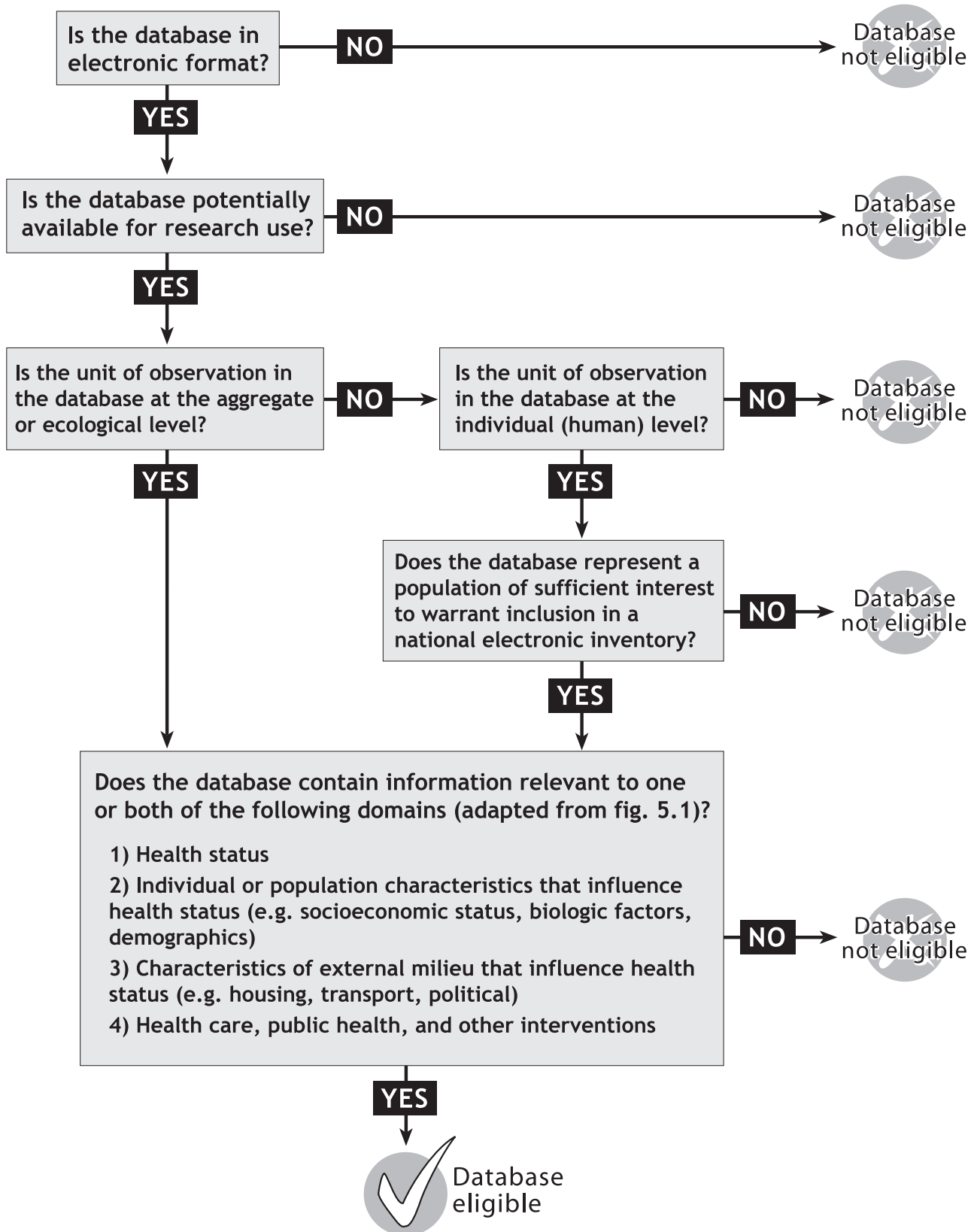
We have developed a first version of a data collection tool that can provide the basis for developing a population health and health services research inventory of databases. To develop this tool, we identified the best ways to collect generic information about databases from a number of existing resources,

and we developed a conceptual model to support consistent recording of information about the content of data sets in a manner that provides relevant information about the population/public health and health services research landscape. Finally, two of the authors (KM and CB) tested the utility of the tool and developed a prototype database by inputting five data sources from a diverse sample of candidate data sets. The results of this data collection exercise—in effect, information that would be provided in an inventory for each of the data sets reviewed—are available in Appendix G.

In general, the data collection tool worked adequately for documenting data holdings relevant for population/public health and health services research and only minor refinements were made during the pilot testing. The pilot test identified that the length of the instrument, the complexity of the database being reviewed, and the reviewer's level of familiarity with the database being reviewed were all associated with time required for completion. In comparison to many existing data inventories, the tool collects an extensive amount of information about each database. Inputting information required approximately 1.5 to 2 hours per database. Since we chose databases with which we had some familiarity, this experience likely underestimates the amount of time that would be required to input data for a larger set of data holdings. Much of the length of the instrument was related to the extensive and consistent documentation of the content of databases.

It is therefore likely that the instrument would have to be refined in the implementation of an electronic database. We believe that the tool will require additional testing and modification before significant investments are to be made in further development, including development of an electronic interface. It will be important to test the utility of information collected with intended audiences. Additional refinements might also be made in relation to levels of investment and stewardship options for further

FIGURE 5.3: DECISION TOOL FOR SELECTION OF RELEVANT DATABASES FOR INCLUSION IN A NATIONAL ELECTRONIC INVENTORY



development. Prior to addressing these issues, it is premature to develop an electronic template. At the same time, given Canada's leadership in using data and research in developing programs and policy, and strategies required for the population health and health services research communities to be successful, further development of this resource may be an important area to pursue.

5.4.2 Potential models for future development of an inventory of databases

Our environmental scan of international and Canadian activities and inventories (see Chapter 4) provided important perspective for our experience with developing and testing the data collection tool. During this process, we identified three major areas of consideration that will need to be addressed by partner organizations in order to move further with implementation. The first is the *model*—what is the nature of the inventory, how often will it be updated, and so on; the second is *stewardship and management*—who will assume responsibility for building and maintaining such an inventory?; and the third is *funding*—from what source(s) will funding for both start-up and ongoing operating costs be derived?

Our review of existing inventories suggested a range of options for developing an inventory of population health and health services research databases, each of which builds in important components. Decisions about implementation revolve around several considerations, including audience targeting, level of detail to be collected, ongoing commitment, cost and stewardship. We describe below four possible models that we have identified as options for future development of an inventory (see also Table 5.1).

Model 1: Framework supports development of one-time cross-sectional inventory

The first option would be to use the framework developed for this project (i.e. the data collection tool and the tool to identify scope of databases for inclusion in an inventory) to develop an inventory on a one-time basis. An example of this model is provided by the inventory of data sets that was prepared for the Health Services Utilization and Outcomes Commission (Alibhai, McCaffrey, and Saunders 2002) in Alberta in a report that documented existing resources at a given point in time.⁵⁷ This would be the least expensive model to undertake, but is also likely to be the least useful to researchers, as it would quickly become out of date and would likely be difficult for new researchers to find unless it was posted on a web portal that is frequently accessed by the population health and health services research community.

Model 2: Centralized framework, with inventory distributed across organizations

A second option would be to develop an inventory on a one-time basis, as outlined above, but to encourage data custodians to provide updated information to an inventory. This might be accomplished by providing data custodians with the data collection tool used to create the inventory, together with training on how to use it. Given some encouragement, if the tool is seen to be useful, custodians may choose to implement it on their own, thereby increasing researcher access to information about potential data resources, albeit in a decentralized fashion. This would also be a relatively inexpensive model to implement, but it is unlikely that data custodians would consistently use a tool developed to map content of data holdings in relation to population health and health services researchers' needs.

⁵⁷In the report, the authors recommended that the project be repeated in two years time to update the information provided. This would allow organizations such as the Regional Health Authorities to respond in a more comprehensive manner and it would also provide opportunities for organizations to identify and report on more 'obscure' data sets in their holdings.

Table 5.1 – Potential models for inventories of data held by multiple custodians*

Features / focus of model	Model 1: One time cross-sectional inventory	Model 2: Centralized framework, with distributed inventory across organizations	Model 3: Ongoing centralized framework and inventory	Model 4: Ongoing centralized framework, inventory and data archive
Primary audience	Funders / researchers	Custodians / researchers	Researchers / funders	Researchers
Expected outcomes/results	Strategic investments	Increased consistency of data documentation	Increased knowledge of available data sets; strategic investment	Increased consistency of data documentation, increased knowledge of available data sets, AND increased access to and use of data
Static vs. ongoing	Static	Mix – dependent on data owners	Ongoing	Ongoing
Electronic /searchable?	Depends on whether publication is paper or electronic	Yes	Yes	Yes
Passive vs. active implementation	No implementation	Passive	Active	Active
Adaptable to evolving requirements?	N/A	Yes – but dependent on custodians for adoption	Yes	Yes
Level of detail accommodated	Accommodated, but cost and size of document / inventory will increase with level of detail. ?? relevance of detail to primary audience	Accommodated, but dependent on custodians to populate the fields	Accommodated, but cost and size of inventory will increase with level of detail	Accommodated, but cost and size of inventory will increase with level of detail
Primary responsibility for ownership //maintenance	N/A – one-time effort	Ownership only required for standards – could be a consortium	Would need to be determined	Would need to be determined
Cost	Minimal – one-time effort	One-time investment with minimal to no ongoing costs	One-time investment with moderate ongoing costs, depending in frequency of updates and new additions	One-time investment with significant ongoing costs
Example	Inventory of data sets in Alberta conducted by CHORUS	No current example	DoCDat.org (UK Directory of Clinical Databases)	U of Michigan ICPSR; UK Data Archive)

Model 3: Ongoing centralized framework and inventory

The third option that we identified would be to develop the inventory and to actively maintain it. An example of this approach is the UK Directory of Clinical Databases (DoCDat). This resource, which was developed to serve the needs of researchers and custodians, has a number of advantages. It has the advantage of being more (potentially) useful to researchers, but comes with higher costs and the additional challenge of identifying an appropriate steward, stewardship model and funding stream. It would also provide opportunities for ongoing development and enhancement of the tool to ensure its utility and relevance to the research community, as well as providing a location for development of a set of other resources to support researchers who are conducting population health and health services database research.

Model 4: Ongoing centralized framework, inventory and data archive

Finally, there is an option to build an inventory that can also serve as an access point to an actual data archive, as seen with many international resources that have been developed to support the research community. The advantage to this model is the support it could provide to the research community, given what we heard in our interviews about the many challenges of access to data. The challenges, however, would be the same as for the third model, and with additional complexities of current privacy legislation governing access to research data sets. It is also unlikely that many data sets used by the health services research community would be available for archiving.

This project was initially given the task of developing an electronic framework for the creation of an inventory of databases relevant to population health and health services research in Canada. We were able to develop a framework that might serve as the content infrastructure for such an endeavour, but our

review of the literature and national and international resources suggests that actually building a prototype “inventory” would be quite premature. In addition to the issues of determining the appropriate model, funding and stewardship, potential funders of an inventory must also consider how this effort fits with other work that is currently underway. International efforts in particular show the potential beyond building a basic inventory of data sets, efforts that will preserve investments in research data and ultimately enhance our understanding of health and the factors that determine it.

Bibliography

This bibliography includes cited references, and additional sources of background material used in the preparation of this chapter.

- Adam, LA, CE Begley, DR Lairson and CH Slater. 1993. *Evaluating the Medical Care System: Effectiveness, Efficiency, and Equity*. Ann Arbor, MI: Health Administration Press.
- Alibhai, A, L McCaffrey and LD Saunders. 2002. *An Inventory of Data Sets in Alberta*. Edmonton, AB: Centre for Health Outcomes Research and Utilization Studies.
- Anonymous. 1999. *A Network for Health Surveillance in Canada*. Alberta Heritage Foundation for Medical Research.
- Bargmeyer BE and DW Gillman DW. 2004. Metadata standards and metadata registries: an overview. Washington D.C., Bureau of Labour Statistics.
- Ben Shlomo Y and D Kuh. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 31(2): 285-293.
- Black, C. 1998. Building a National Health Information Network. In *Canada Health Action: Building on the Legacy, papers commissioned by the National Forum on Health*. Sainte-Foy, Quebec: Editions MultiMondes.
- Black, N. et al. 2003. Directory of clinical databases: improving and promoting their use. *Qual Saf Health Care*. 12: 348-352.
- Black, N. 2004. Interview by Kim McGrail. 16 July.
- Choi, B. 2000. Perspectives on Epidemiologic Surveillance in the 21st Century. *Chronic Diseases in Canada*. 19(4): http://www.hc-sd.gc.ca/pphb-dgspsp/publicat/cdic-mcc19-4/b_e.html
- CIHR: Canadian Institutes of Health Research. Institutes of Population and Public Health and Health Services and Policy

- Research. 2002. *Population-based health and health service data in Canada: current status and future health research potential*. Ottawa: CIHR.
- CIHR: Canadian Institutes of Health Research. Institute of Population and Public Health. 2002. *Mapping and tapping the well-springs of health: strategic plan 2002-2007 – why “population health?”*. Cat. No. MR21-36/2002E. 2002. Ottawa, ON, Canadian Institutes of Health Research.
- Deerfield Group. 2002. *Identifying Health Infrastructure Priorities of Canadian Health Services, Clinical and Population Health Researchers: Beginning the Dialogue*. Workshop Report. Burlington, Ontario: The Deerfield Group
- Donabedian, A. 1966. Evaluating the quality of medical care. *Milbank Mem Fund Q* 44(3):Suppl:166-206.
- Evans, RG, and GL Stoddart. 1990. Producing health, consuming health care. *Soc Sci Med* 31(12):1347–1363.
- Frank JW and JF Mustard. 1994. The determinants of health from a historical perspective. *Daedalus* 123(4): 1-17.
- Frank JW. 1995. Why “population health”? *Can J Public Health*. 86(3): 162-164.
- Frank, JW. 1995. The determinants of health: a new synthesis. *Curr Issues Public Health* 1:233-240.
- Goel, V. 1998. Health services research in breast cancer: background paper for a Canadian Breast Cancer Research Initiative workshop. *Cancer Prevention & Control*. 2(4): 173-178.
- Health Canada. 2002. *Towards a common understanding: clarifying the core concepts of population health: a discussion paper*. Cat. No. H39-391/1996E. 2002. Ottawa, ON, Health Canada.
- Hertzman C and M Wiens. 1994. *Surveys and longitudinal research in Canada - a source document of relevant studies*. CIAR Programs in Human Development and Population Health: The Child Longitudinal Studies Project. Working paper no. 3. Toronto, ON, Canadian Institute for Advanced Research.
- Hillemeier, MM, J Lynch, S Harper and M Casper. 2003. Measuring contextual characteristics for community health. *Health Services Research* 38(6 Pt 2):1645-717.
- Holman, CD et al. 1999. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Australian and New Zealand Journal of Public Health*. 23(5): 453-459.
- LaMontagne AD et al. 2002. Exposure databases and exposure surveillance: promise and practice. *AIHA Journal: a Journal for the Science of Occupational & Environmental Health & Safety*. 63(2): 205-212.
- Linger, H et al. 1998. Conceptual development of an information systems framework for improving continuity in epidemiological research. *European Journal of Epidemiology*. 14: 587-593.
- McGlynn EA. 1998. *Health information systems : design issues and analytic applications*. Santa Monica, CA: Rand.
- Melnychuk, D, Y Moride and L Abenhaim L. 1993. Monitoring of Drug Utilization in Public Health Surveillance Activities: A Conceptual Framework. *Canadian Journal of Public Health*. Jan.-Feb.84(1):.45-49.
- Meriwether, R. 1996. Blueprint for a National Public Health Surveillance System for the 21st Century. *Journal of Public Health Management Practice*. 2(4): 16-23.
- National Committee on Vital and Health Statistics. 2002. Shaping a health statistics vision for the 21st century: final report. Washington, DC, National Committee on Vital and Health Statistics.
- Roos, LL, RA Soodeen, R Bond and C Burchill. 2003. Working More Productively: Tools for Administrative Data. *HSR: Health Services Research*. Oct 38(5): 1339-1357.
- Roos, NP et al. 1995. A population-based health information system. *Med Care* 33(12 Suppl):DS13–DS20.
- Shatin D. 2001. Organizational context and taxonomy of health care databases. *Pharmacoepidemiology & Drug Safety* 10(5): 367-371.
- Slater, Carl H. 1999. Outcomes Research and Community Health Information Systems. *Journal of Medical Systems*. 23(4): 335-347.
- Studnicki J, SL Luther SL, J Kromrey J and B Myers. 2001. A minimum data set and empirical model for population health status assessment. *Am J Prev Med*. 20(1): 40-49.
- Wannag A. 1993. Classification system for work in occupational health services. *Scandinavian Journal of Work, Environment & Health*. 19(6): 390-393.

Chapter 6: Next Steps and Recommendations

This project was designed and funded with the aim of enhancing researchers' access to and use of current data infrastructure and in making wise additional investments in data and analytic capacity. Our objectives in undertaking this project were to identify the barriers that currently stand in the way of that access and use, to test assumptions about the utility of developing a national inventory of population-based research databases, and to structure recommendations on strategic investments and directions for the future.

This work was based on a review of literature and resources related to privacy, access, data inventories, data archives, and conceptual frameworks and taxonomies, and a series of interviews with collectors, custodians and users of data. The literature review and interviews offer a common view of support for research conducted in the public interest with an ultimate goal of improving the population's health. Differences in perspective, however, about how best to make this happen point to tensions and barriers in the current Canadian research landscape. Regardless of these barriers, this common starting point—of support for privacy-sensitive research conducted in the public interest—should be recognized.

Beyond this basic starting point, there is a challenge in seeing how Canada can most effectively move forward. Canada has an international reputation based on the development and implementation of a population health framework—an understanding and recognition of the many factors that influence the health status of individuals and populations. Canada also is known for the collection and research use of administrative data related to health care services. In part, this reputation is based on the availability of universal and comprehensive data about the use of health care services, data that exist because of the funding and administrative structures of provincial, territorial and federal health care services. This reputation also comes from recognition of Canadian researchers as innovators in understanding the power that such data hold, and in converting that understanding into

research findings that have provided a wealth of evidence for the policy development process (see, for example, Roos et al. 1995).

Our work suggests, however, that Canada is not currently recognized as a leader when it comes to the systematic organization, archiving, documentation of, and access to data relevant to population health and health services research. The recommendations that follow highlight opportunities to change this, and to push our national efforts forward in support and recognition of the excellent work already done and the excellent work that can and should follow.

6.1 Protecting privacy

Recommendation 1:

CIHR should take a lead in coordinating a series of activities to address privacy issues that are specific to the population health and health services research community. This work includes:

- a) Clarifying the definition of research that has “public value”;
- b) Developing a constellation of privacy tools and techniques (including best practice guidelines) to assist researchers and data custodians in protecting privacy while allowing access to data;
- c) Strengthening the role of research ethics boards, increasing and harmonizing expertise;
- d) Influencing the development and interpretation of regulatory and legislative frameworks to ensure they support privacy-sensitive research, and where possible, that they are harmonized across jurisdictions; and
- e) Engaging with the public about the value of health and health services research and how it should be conducted in a privacy-sensitive manner.

The landscape has shifted substantially since the funders released this original RFP in 2002. The *Personal Information Protection and Electronic Documents Act (PIPEDA)* has come into force,

several provinces have introduced or modified privacy legislation and Canada Health Infoway has begun infrastructure development in support of an electronic health record, to name just a few changes. Meanwhile, the funders of this project have continued with their own initiatives; funding research, building data and resource inventories and conducting research. The research context is changing rapidly. In order for strong research to continue, the research community—including data collectors, data custodians, research ethics boards, and researchers themselves—requires a common understanding of each party’s roles and responsibilities. In addition to this, there must be a common understanding of the potential of research to contribute to health and well-being, the implications of the legislation that governs the research process, the mandates and actions of data collectors and custodians and the extent of available funding.

There has been a lot of work in this area over the last few years, which CIHR and other research funders could use as a springboard for making changes. For example, CIHR recently released a consultation draft on best practices in health research (CIHR 2004), a workshop series sponsored by CIHR looked at the possibility of creating some harmonization in research practices (Slaughter et al. 2004), and a project is currently underway to assess the practices of research ethics boards across the country.⁵⁸ These efforts and recent work of the General Accounting Office and National Center for Health Statistics in the United States offer several suggestions for supporting the research community, including developing privacy tool kits to assist research organizations and data custodians in protecting privacy while allowing access to data; developing best practice guidelines for conducting research in the areas of population health and health services research; developing models and best practices in data stewardship/custodianship; and developing material and other supports to harmonize the work of research ethics boards across the country.

Each of these areas requires further work. More importantly, all of these strands need to be considered as contributing to a greater whole, which is the development and support of population health and health services research in a manner that honours a commitment to protection of privacy. For this coordination to occur, it needs to be supported and nurtured by a national body such as CIHR. The ultimate goal is for research to be conducted in a system that is sufficiently constructed and governed such that it is difficult to make a mistake; a system in which all parties understand their roles and responsibilities, and in which protecting privacy while promoting solid, public value research is central.

It is not sufficient, of course, for the research community to act as an insular entity. We must find ways to engage with the public about the value of health and health services research. The research community must also work with the public, as well as privacy experts, to find solutions for conducting research in a privacy-sensitive manner. CIHR issued a request for proposals in this area in 2003, and there are currently four projects funded through this initiative. Results of these projects should be considered part of this coordinated effort.

6.2 Improving and increasing access to data

Recommendation 2:

CIHR should convene and lead a “coordinating body” that will focus on improving access to population health and health services research data and that will be charged with reviewing and carrying forward the recommendations in this report.

Recommendation 3:

CIHR, as the lead organization for health research in the country, and in cooperation with other funders of health research, should strongly encourage key national and provincial data

⁵⁸A description of this project by Willison and colleagues can be found at www.cihr.ca.

custodians to review their mandates, with the goal of clarifying and increasing their commitment to providing data and other supports for population health and health services research.

Recommendation 4:

Data custodians of population health and health services data, including the Canadian Institute for Health Information and provincial data custodians, should be encouraged to work with privacy experts and the research community to create and make available public use microdata sets as well as to provide access to more detailed microdata sets for publicly funded research.

Recommendation 5:

Provincial and regional custodians of population health and health services data should develop clear processes and equitable costing mechanisms for making data available to researchers.

Recommendation 6:

CIHR should support the costs of conducting data-based research in population health and health services research by:

- a) Under certain circumstances, allowing operational research budgets to include the costs of archiving and documenting large-scale data collection efforts, where there is intent to make those data more broadly available to the research community;
- b) Developing funding streams that parallel the “equipment grant” program used by the basic and clinical health research domains.

Recommendation 7:

CIHR should actively pursue opportunities to work with current initiatives with the potential to improve access to research data that supports development of population health and health services research:

- a) In the ongoing National Consultation on Access to Scientific Research Data, to ensure that the special circumstances

around access to population health and health services data (i.e. privacy considerations around personal health information and dependence on non-research data collectors and custodians) are addressed;

- b) In influencing Canada Health Infoway, to explicitly consider and build in mechanisms to support researcher access to data as it invests in prototypical development of information infrastructure.

Recommendation 8:

CIHR should work with partners to develop a web-based “population and public health and health services research” portal that could house an electronic inventory as well as related tools to support the research community to use existing data resources efficiently and in a privacy-sensitive manner.

Recommendation 2: Creation of a coordinating body

This project was funded by a consortium of data, statistical and funding organizations with an interest in supporting population health and health services research. There are many players in this field, with differentiated and overlapping roles and responsibilities. If there is a sincere interest in providing more and better supports to the constituent research communities, it is not clear how this can happen without some coordinating body or committee to plan and oversee the work.

This is not a recommendation for a new pan-Canadian or national agency. Instead, we suggest that CIHR, as the primary funder of this work and the largest funder of health research in Canada, take the lead on formally convening a group of individuals representing the various stakeholders in population health and health services research. In the first instance, this group would require only a small budget to support consideration of the recommendations in this report. In the longer term, the group would need to take responsibility for identifying and securing

whatever funding is necessary to undertake the tasks they prioritize.

The coordinating body should be viewed as a semi-permanent structure, as the recommendations outlined here will clearly take years rather than months to complete. In addition, there will always be new and evolving issues that are relevant to all stakeholders with respect to this growing and important area of research.

Recommendation 3: Review of organizational mandates

Data collectors and custodians expressed hesitancy in allowing access to data holdings for several different reasons. In some cases, there was a belief that support for researchers was not consistent with organizational goals or mandates; in others there was a concern that there was insufficient funding to support the research community; and in others, there was trepidation that the data would be misinterpreted—not wilfully, but because complex survey and administrative data demand a sophisticated understanding to ensure appropriate use.

National agencies including Statistics Canada, Health Canada, CIHR, and CIHI need to examine their roles and mandates with respect to the provision of access to data. If research that has the potential to improve the population’s health is a key goal of these agencies, then this must be mirrored in the mandates of agencies responsible for safeguarding and providing data. This issue is particularly noteworthy with respect to Canada Health Infoway, an organization established in 2002 without a mandate to connect to and support the research community. Given the remarkable potential of the electronic health record for research purposes—albeit with attendant privacy challenges—this remains a gaping hole.

The same is true for provincial ministries and agencies, with the additional challenge that there is a less consistent tradition of these organizations

working with the research community. There is a great deal of potentially useful data that reside outside national agencies, most notably in provincial departments and ministries charged with operating programs and services for the public. The primary mandates of these organizations are and will remain the operation of those services, but there is much that could be done to improve the relationship between these agencies and researchers, particularly researchers external to those organizations. This would be consistent with the tremendous shift over the last decade or more to a recognition of the importance of making public policy decisions based on, or at least informed by, evidence.

It is a small but important, and often neglected, step to move from data collection for operational purposes to viewing those same data as a powerful potential resource for research. There is a “sunk cost” associated with collecting data to support operating public programs. These organizations should be encouraged to recognize that the costs associated with supporting researchers in their use of data are minimal compared to that sunk cost, and will almost certainly be far outweighed by the benefit (in evidence) gained from the research.

Recommendation 4: Creating better access to more data

Statistics Canada has done a great deal of work to develop public use microdata files that are available (through the Data Liberation Initiative) to university-based researchers in Canada free of charge and without access protocols. The looser—though not absent—control mechanisms for these data sets are possible because they have been reviewed, modified and “cleaned” to remove any practical possibility of identification of individuals. Once identification of individuals is no longer possible, the privacy restrictions on the use of these data no longer apply.

It would be useful to develop this model further for other data sources. There are developmental costs

involved in such an undertaking, but the payoffs in improved access and increased use of data would very likely far outweigh these costs. The expertise at Statistics Canada in this area and the agency's reputation for data protection would be useful for such a development process, and for encouraging the cooperation of wary data custodians. This is particularly true for CIHI, which should be encouraged to develop more freely available, unencumbered data sets along the lines of the national data sets provided through the Healthcare Cost and Utilization Program in the United States.

Beyond the production of public use data files, data collectors/custodians should consider allowing greater access for researchers to microdata, under controlled, privacy-sensitive circumstances. With the privacy tools in place, and a refinement of mandates to reflect the importance of research, this should be an achievable goal.

Recommendation 5: Making the data access process clear

Data custodians are responsible for protecting the confidentiality of data and ensuring that they and all parties to whom they disclose data meet the requirements of privacy legislation and other governing policies. This is an important role, but one that can be balanced with the interests and needs of researchers, if we can make progress on the recommendations above. In the interim, and at the very least, data collectors and custodians should make clear by what process, under what circumstances and with what cost they will make data available to researchers.

Recommendation 6: Developing explicit funding mechanisms

The data archiving policies of the UK's Economic and Social Research Council and Medical Research Council, and the Social Sciences and Humanities Research Council of Canada, provide a useful template that should be considered by CIHR. It is not

the case that every funded research project would produce a data set that could or should be archived for future use. But there clearly are cases where a data archiving policy would serve the interests of CIHR and the population health and health services research communities. The proposed Canadian Longitudinal Study on Aging, for example, is exactly the kind of survey-based study that other funding agencies are starting to insist become publicly archived and available in some form. This archiving function should include all aspects of documenting and storing data, to ensure adequate oversight and protection of the data and the detailed information about the data required for subsequent analyses.

In addition, CIHR should also give some consideration to making infrastructure type funding available, in parallel to the "equipment grant" model used by clinical and bench scientists. CIHR could, for example, fund the purchase of computer servers, necessary for the storage and safe-keeping of larger data sets common in analyses of administrative data.

Recommendation 7: Pursuing the potential of current initiatives

CIHR is participating in the National Consultation on Access to Scientific Research Data. It is recommended that CIHR bring forward to this process consideration of the "special" nature of the majority of population health and health services research data. For example, in contrast to data about weather patterns or air quality, data in health services and population health often (or usually) relate to individuals, thereby raising privacy concerns. In addition, researchers in these fields are often dependent on non-researcher data custodians (such as ministries of health or education) for access to data. These agencies may not have considered the utility of the data they hold, much less the importance of archiving the data for future use. Results of the National Consultation have the potential to help inform these agencies of the treasure they hold, and its research potential.

Consideration should also be given to the opportunity to influence models for ensuring research access to new data and information development that is being supported by Canada Health Infoway, especially as to date, researcher interests have not been considered in the development of the electronic health record in Canada.

Recommendation 8: Creating a research portal

Interviews and discussions with key informants suggested that an inventory of research databases is only one piece of a larger set of investments that must be developed to support the population health and health services research communities. It was suggested that development of a web-based portal that contains other resources for researchers, such as privacy training tools, modules providing information about best practices for data documentation and analysis, and access to related websites, including those of relevant data providers, would add to the value of an electronic data inventory. This could be viewed as a “first step” and a logical home for an database inventory, if one is to be developed.

6.3 Developing an inventory of population databases

Recommendation 9:

The partners should review the findings from the interviews and the survey of existing activities to reassess their commitment to building, maintaining and refining an inventory.

Recommendation 10:

If the partners wish to proceed with development of an inventory, they should develop an appropriate vision and business plan. This vision/business plan should:

- g) define the objectives of the resource;
- h) identify the primary customers to be served;
- i) identify a model that can build on Canadian activities already underway

to document agency- and topic-specific data holdings;

- j) identify a steward or host agency that can competently develop and manage the resource;
- k) identify ongoing funding to support development over a period of at least five years; and
- l) identify an evaluative process to ensure the resource developed meets the needs of all relevant stakeholders

Recommendation 9: Reassessing commitment

Clear consideration needs to be given to specifying a model for future development before further work is undertaken to develop an electronic inventory. Given the lack of a majority of support from researchers interviewed for this project, we believe it is important for the partners to reconsider their commitment to developing an inventory, especially in light of the ongoing requirements for maintenance, updating and refinement necessary for a truly useful inventory. Our review of existing activities suggested many resources, such as portals targeted at the research community, that might serve a more useful role in the short term. At the same time, we suggest that consideration be given to the merit of pursuing the development of a data inventory if there is no concurrent, or long term, goal of providing, or assisting in the provision, of access to the data sets that are catalogued.

Recommendation 10: Vision and business plan

Across all of the potential inventory models outlined in Chapter 5, there is considerable latitude to develop the resource in different ways to respond to the needs of different audiences. At some level, there will need to be consensus about what the resource should be designed to do. Should it serve as a tool to simply make people aware of existing data, give them a general sense about what elements are included and provide contact information for

custodians? Or should it be a much richer and more interactive resource? For example, researchers who are new to the area of population health and health services research will have different information needs than those who are already established in this area. The former will likely require information that can help them learn about data resources in the area they wish to study; they will likely also require additional resources to help them use existing data more proficiently. The latter group will likely require highly detailed information provided in an interactive format. Ideally, a resource that could serve a range of research needs across the spectrum of researchers working in pillars 3 and 4 could be developed. This process will require identification of a steward who can interact with users to develop and improve a resource over time and provision of ongoing funding to support development over a number of years. The funding partners of the current project may have underestimated the complexity of developing a resource that will be helpful to the research community.

Our discussions with key informants suggested that there is no consensus about how to respond to the issues outlined above. All agreed that there is no single agency that has a mandate to take on the development of an inventory of databases, but it was suggested that because the resource is intended primarily to support the research community, CIHR should serve as the lead in developing the vision and business plan, in partnership with other agencies (including Statistics Canada, CIHI, Health Canada, the Public Health Agency of Canada, the Advisory Committee on Health Infrastructure (ACHI) and others). A number of different possibilities were identified for a steward or host agency, including CIHI, Health Canada or a university-based group with experience in database inventory work.

A decision about which model to support must be tied to the availability of funding and identification of an appropriate steward for further development of the

resource. Choice of a model has implications for the initial expense of development and ongoing costs for maintenance, but also relates to the likelihood that the resource will meet all potential user needs. We believe that it does not make sense to proceed with Models 1 or 2 identified in our report, and that Model 4 is impractical with current levels of investment (as outlined in Chapter 5). An “information-only” resource probably makes sense for the current landscape, given the distributed nature of data. However for future development, consideration should be given to opportunities to build research infrastructure that will include data archives and provide differential access for system management and research.

6.4 Concluding remarks

There is a great deal that could be done to support the community of population and public health and health services researchers in Canada. Building an inventory of population-based databases, as envisioned by the funders of the RFP for this project, is one option. But there are many issues to consider prior to starting down that particular path.

One of the most compelling things we discovered in the course of this project is the sheer number of players involved in conducting research, collecting or holding data relevant to population health and health services research, commenting on research and the research process, or working on developmental projects where there is a clear crossover of interest. Without some form of coordination among all these parties, efforts will be duplicated, or worse yet, we will lose opportunities purely through not knowing that they exist.

There is a clear role for a body, working group, or some other organization to take the recommendations in this report and coordinate or monitor activities relevant to them. At the very least, a body of this sort would need to include representatives from CIHR, CIHI, Statistics Canada and Health Canada,

as well as provincial departments and agencies, privacy experts, public advocates, and of course, researchers.

Less clear is how such a group might be formed and maintained. Our hope is that CIHR will recognize the critical importance of this work in supporting its researchers and will take on this daunting but important challenge.

References

CIHR: Canadian Institutes of Health Research. 2004. *Guidelines for Protecting Privacy and Confidentiality in the Design, Conduct and Evaluation of Health Research. Best Practices. Consultation Draft*. Ottawa: CIHR.

Roos, NP et al. 1995. A population-based health information system. *Med Care* 33(12 Suppl):DS13–DS20.

Slaughter, P et al. 2004. *Harmonizing Research and Privacy: Standards for a Collaborative Future. Final Workshop Summary*. Toronto: Institute for Clinical Evaluative Sciences.

APPENDIX A: Mission and Mandates of Canadian and international data and research organizations

Table 1: Mission and Mandates of Canadian data and research organizations

Organization	Mission Statement	Mandate
FEDERAL GOVERNMENT		
HEALTH CANADA (HC)	HEALTH CANADA IS RESPONSIBLE FOR HELPING THE PEOPLE OF CANADA MAINTAIN AND IMPROVE THEIR HEALTH.	<i>Provides national leadership to develop health policy, enforce health regulations, promote disease prevention and enhance healthy living for all Canadians.</i>
Information, Analysis and Connectivity Branch (IABC) in HC	Facilitate evidence-based decision making at all levels of the health system.	<i>Works to improve the analytic basis of decision making, develop the creative use of the information highway in the health sector, and in cooperation with the provinces and territories, the private sector and international partners, provide advice, expertise and assistance with respect to information management and information technology, planning and operations.</i>
Applied Research and Analysis Directorate (ARAD), in IABC, HC	ARAD helps to build the analytical foundation for Health Canada's policy decision making and performance measurement and reporting.	<i>ARAD conducts economic analysis of health policy issues, funds external policy research in selected policy areas, and runs a policy research publications program. ARAD develops the department's program evaluation plan and standards, as well as conducting evaluations of key departmental policies and programs. ARAD develops federal policy on federal investments in Canada's health statistics system and coordinates department core data requirements with data providers. ARAD has also developed a data dissemination system to give Health Canada analysts single-tool access to core data from a wide variety of sources.</i>
Research Management and Dissemination Division (RMDD) in ARAD, IABC, HC	Improve the availability, quality and use of evidence in decision making within Health Canada.	<i>Identifies future policy research needs, funds research and communicates the results within Health Canada and externally.</i>

<p>Centre for Surveillance Coordination (CSC), Public Health Agency of Canada</p>		<p><i>To collaborate with IT professionals, health professionals and policy makers on the development, maintenance and use of health information, tools and skills to enable timely and informed decision making for public health that results in improved public health policies, programs and interventions that protect and promote the health of Canadians.</i></p>
<p>Centre for Chronic Disease Prevention and Control (CCDPC), Public Health Agency of Canada</p>	<p>Works with stakeholders at all levels to provide pan-Canadian and international leadership in chronic disease prevention and control through integrated policy and program development, surveillance and knowledge development and dissemination.</p>	
<p>CANADA HEALTH INFOWAY</p>	<p>Fostering and accelerating the development and adoption of electronic health information systems with compatible standards and communications technologies on a pan-Canadian basis with tangible benefits to Canadians.</p>	
<p>CANADIAN HEALTH SERVICES RESEARCH FOUNDATION (CHSRF)</p>	<p>To support evidence-based decision making in the organization, management and delivery of health services through funding research, building capacity and transferring knowledge.</p>	<p><i>To enhance the quality and quantity of research that responds to the needs of health system decision makers.</i></p> <p><i>To get needed research into the hands of health-system managers and policy makers in the right format, at the right time, through the right channels.</i></p> <p><i>To help health system managers, policymakers and their organizations to routinely acquire, appraise, adopt and apply relevant research in their work.</i></p> <p><i>To bring researchers and decision makers together regularly to understand each other's goals and professional culture, influence each others work and forge new partnerships.</i></p>
<p>CANADIAN INSTITUTE FOR HEALTH INFORMATION (CIHI)</p>	<p>Working to improve the health of Canadians and the health care system by providing quality, reliable and timely health information.</p>	<p><i>To coordinate the development and maintenance of a comprehensive and integrated approach to health information for Canada and to provide and coordinate the provision of accurate and timely data and information require for establishing sound health policy, effectively managing the Canadian health system and generating public awareness about factors affecting good health.</i></p>

<p>Canadian Population Health Initiative (CPHI) in CIHI</p>	<p>To foster a better understanding of factors that affect the health of individuals and communities and to contribute to developing policies that reduce inequities and improve the health and well-being of Canadians.</p>	<p><i>Provides analysis of Canadian and international population health evidence to inform policies that improve the health of Canadians.</i> <i>Funds research and builds research partnerships to enhance understanding of research findings and to promote analysis of strategies that improve population health.</i> <i>Synthesizes evidence about policy experiences, analyzes evidence on the effectiveness of policy initiatives and develops policy options.</i> <i>Works to improve public knowledge and understanding of the determinants that affect individual and community health and well-being.</i> <i>Works within the CIHI to contribute to improvements in Canada's health system and the health of Canadians.</i></p>
<p>CANADIAN INSTITUTE OF HEALTH RESEARCH (CIHR)</p>	<p>To excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system.</p>	<p><i>Canada's major federal funding agency to health research.</i></p>
<p>Institute of Health Services and Policy Research, CIHR</p>	<p>Dedicated to supporting innovative research, capacity-building and knowledge translation initiatives designed to improve the way health care services are organized, regulated, managed, financed, paid for, used and delivered in the interest of improving the health and quality of life of all Canadians.</p>	
<p>Institute of Population and Public Health, CIHR</p>	<p>To support research into the complex interactions (biological, social, cultural, environmental) which determine the health of individuals, communities, and global populations; and into the application of that knowledge to improve the health of both populations and individuals.</p>	

<p>NATIONAL RESEARCH COUNCIL OF CANADA (NRC)</p>		<p><i>Undertaking , assisting or promoting scientific industrial research in different fields of importance to Canada.</i></p> <p><i>Establishing, operating and maintaining a national science library.</i></p> <p><i>Publishing and selling or otherwise distributing such scientific and technical information as the Council deems necessary.</i></p> <p><i>Investigating standards and methods of measurement.</i></p> <p><i>Working on the standardization and certification of scientific and technical apparatus and instruments and materials used or usable by Canadian industry.</i></p> <p><i>Operating and administering any astronomical observatories established or maintained by the Government of Canada.</i></p> <p><i>Administering research and development activities including grants and contributions used to support a number of international activities.</i></p> <p><i>Providing vital scientific and technological services to the research and industrial communities.</i></p>
<p>SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL OF CANADA (SSHRC)</p>		<p><i>Arms-length federal agency that promotes and supports university-based research and training in the social sciences and humanities.</i></p>
<p>STATISTICS CANADA</p>	<p><i>Produces statistics that help Canadians better understand their country – its population, resources, economy, society and culture.</i></p>	<p><i>To collect, compile, analyse, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and conditions of the people of Canada.</i></p> <p><i>To collaborate with departments of government in the collection, compilation and publication of statistical information.</i></p> <p><i>To take the census of the population and the census of agriculture.</i></p> <p><i>To promote the avoidance of duplication in the information collected by departments of government.</i></p> <p><i>To promote and develop integrated social and economic statistics pertaining to the whole of Canada and to each of the provinces.</i></p>

Table 2: Mission and Mandates of US data and research organizations

Organization	Mission	Mandate
<p>CENTRE FOR DISEASE CONTROL AND PREVENTION (CDC)</p>	<p>Lead federal agency for protecting the health and safety of people—at home and abroad, providing credible information to enhance health decisions and promoting health through strong partnerships.</p>	<p><i>National focus for developing and applying disease prevention and control, environmental health and health promotion and education activities designed to improve the health of the people of the United States.</i></p> <p><i>Specific statement about data—by working with public health and grassroots partners, and by leveraging the voices of the internet and communication media, we ensure the best health and safety information is accessible to the communities and people who need it every day.</i></p>
<p>National Centre for Health Statistics, CDC</p>	<p>To provide statistical information that will guide actions and policies to improve the health of the American people.</p>	<p><i>The nation’s principal health statistics agency, NCHS leads the way with accurate, relevant and timely data.</i></p>
<p>NATIONAL INSTITUTES OF HEALTH</p>	<p>Science in pursuit of fundamental knowledge about the nature and behaviour of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability.</p>	<p><i>Foster fundamental creative discoveries, innovative research strategies and their applications as a basis to advance significantly the Nation’s capacity to protect and improve health.</i></p> <p><i>Develop, maintain and renew scientific human and physical resources that will assure the Nation’s capability to prevent disease.</i></p> <p><i>Expand the knowledge base in medical and associated sciences to enhance the Nation’s economic well being and ensure a continued high return on the public investment in research.</i></p> <p><i>Exemplify and promote the highest level of scientific integrity, public accountability and social responsibility in the conduct of science.</i></p>

APPENDIX B: Interview method

To explore these issues, key informant interviews were conducted with collectors, custodians, and users of population-based health and health services databases. Specific objectives of the interviews were:

- 1) To identify the range of issues and challenges faced by researchers when accessing and using existing population-based health and health services research data; and
- 2) To identify the range of issues and challenges faced by population-based health and health services research data collectors and custodians in developing and maintaining databases and in granting data access to researchers.

In addition, interviews were used to inform the other phases of the project; namely:

- 3) To obtain feedback on the utility and development of an electronic inventory of population-based health and health services databases; and

Specifically, the perspectives of collectors/custodians and users were sought to identify relevant and useful approaches for development of an electronic inventory of population-based health and health services databases; i.e. the desirability and utility of such a resource, key capabilities it should have; recommendations on who should have ongoing responsibility for ensuring accuracy, functionality, and maintenance, and finally, the potential challenges, and opportunities of such a resource.

The final objective of the interviews was:

- 4) To identify cross-agency strategic investments for CIHR and its partners for new population-based health and health services research data.

Collectors of population-based health and health services data range from national agencies (e.g. Statistics Canada, Canadian Institute for Health Information), federal and provincial ministries of health, regional authorities, public health units, disease-specific organizations (e.g. cancer agencies), regulatory bodies, and workers' compensation plans. While these various data collectors are custodians of their own data sets, there are also other bodies that have licensing agreements with major data collectors to hold and conduct research on data that are transferred to them (e.g. research centres, such as the Centre for Health Services and Policy Research).

Users of population-based databases range from governments and their agencies, university-based researchers, scientists in data centres or other special purpose agencies, such as disease-specific organizations (e.g. cancer agencies) or service-specific bodies (e.g. wait list registries), regulatory bodies, and professional associations.

There are no clear demarcations in the categorization of data collectors/custodians and data users. There is considerable overlap in functions across these categories. Some data users collect their own data and are custodians of those data. Some data custodians are also data users. However, for the purposes of this study, participants were assigned to either of the two categories based on their major functions.

A list of *major* population-based health and health services data collectors/custodians and users encompassing all

the above categories was drawn up. Special attention was paid to ensure regional representation. For expediency, the list of users only included seasoned researchers/scientists and as a result, introduces a bias in the responses. However, unlike the novice, the seasoned researcher is likely to have a breadth of experience to draw upon.

Participants were contacted by email to inform them of the purpose and nature of the study and to ask for their consent to participate in an interview. Semi-structured, open-ended interview questionnaires were developed and emailed in advance to participants. The interviews were conducted by telephone and simultaneously transcribed on a word processor. Interviews lasted between one to three hours. Interviews continued until regional representation was achieved and saturation was reached on the emerging themes.

In total 43 interviews were conducted; 18 with data users and 25 with data collectors/custodians. Table 1 shows the breakdown by category of participant and by region.

Table 1: Breakdown of respondents by region

Regions	Users	Collectors/ Custodians	Total
Western Canada (BC, AB, SK, MB)	5	9	14
Central Canada (ON, QC)	9	6	15
Eastern Canada (NB, NS, PEI, NF)	4	5	9
Territories			
National (agencies with a national scope of activities)		5	5
Total	18	25	43

APPENDIX C: Interview instruments

Interview schedule for collectors/custodians of population-based health and health services data

Background

We are interested in the views of individuals/organizations who collect data (data collectors), hold data (data custodians), or do both. In particular, we are interested in those who collect population-based health and health services research data and/or hold population-based health and health services research databases, registries, and repositories. By data collectors we mean individuals or organizations that collect primary data on a somewhat routine basis for their own research purposes, such that they hold a data set for a period of time (e.g. a repeated survey of a cohort), or that they systematically input administrative data into a data set (e.g. provincial ministries of health that maintain physician billing information).

Data custodians are individuals/organizations that are responsible for the maintenance and upkeep of data holdings. Data collectors by virtue of their data holdings become data custodians. But data custodians also include individuals/organizations who do not collect the data themselves but through an agreement are provided with a data set which is updated on a regular basis; e.g. independent research institutes that are systematically provided administrative data by provincial ministries of health.

Description of database

1. If you believe it would be helpful to me to read background materials beforehand on the nature and structure of your data holding(s) and rules governing its access and use, if applicable, could you please direct me to the appropriate web-based resource or provide me with the relevant information. We appreciate that you may have a number of databases. In responding to our questions, it may be helpful to know that we are interested in capturing the range of experience data collectors/custodians have in developing, maintaining and updating their databases or in the development and implementation of new systems.

Development/maintenance of data sets

This section of the interview will focus on issues and challenges you face in the development and implementation of new systems, as well as the maintenance and updating of existing data holdings.

1. To begin with, could you please describe your role/responsibilities with respect to your data holdings and the relevant background you bring to your responsibilities?
2. a) What are the major barriers you face in database development, implementation of new technologies, and database maintenance and update? To assist you in thinking about your response, we offer the following list as examples of areas that may contain or create barriers:
 - funding
 - data linkage
 - technical capacity of staff and training
 - development of new technology; e.g. software, hardware or new approaches
 - data quality and completeness, etc.

- b) Where do you think future pressures/areas of concern will be with respect to database development, implementation of new technologies, and database maintenance and update?
3. a) In the current data environment, what is your priority for allocating resources across the areas of data development, data collection, data cleaning, maintenance/upkeep, updating, training, or any other areas that are not listed? That is, where are you putting most of your resources?
- b) Where do you think future pressures will be in the allocation of resources? That is, in which areas do you believe you will need to put more resources?
4. a) Data users frequently express concerns about data quality. With respect to your data holdings, is data quality an issue for you? What are the sources and types of errors usually encountered? Are there strategies you use to enhance data quality of your data holdings?
- b) If you are a data custodian who does not collect the original data, do you have any influence in enhancing data quality? What are the sources and types of errors usually encountered?
5. As we move forward, we are understanding that there are options and opportunities for enhancing data quality. How have you or your organization responded to this either in developing approaches for data quality improvement (as a data collector) or the ability to influence data quality improvement (as a data custodian)?
- b) What are the barriers to improving data quality?
- c) What kind of an investment and/or resources would you need to maintain or improve data quality?

Data access

This section deals with the ability of individuals or organizations to access your data holdings for research purposes done in the public's interest, and some of the challenges you and they face in this process. Once again, you may find it beneficial to me and a better use of your time to direct me to or share with me in advance access protocols/guidelines/regulations for access to your data holdings.

1. Do you make your data holdings available to others for research purposes? If not, what are the reasons for not extending access and are you interested in doing this?
2. a) What types of individuals/organizations are allowed access (or would you be interested in allowing access to, if you currently do not permit access) to your data sets? For what purpose and under what conditions?
- b) If you do allow access to your data holdings, who are the individuals/organizations that have typically accessed the data sets? For what purpose and under what conditions?
- c) Whether you do or do not permit access to your data holdings, are there individuals/organizations that have expressed an interest in having access? If so, what are the reasons for denying access? What would it take to extend access to these individuals/organizations?

- d) Are the types of individuals/organizations allowed access to linked data sets different from those allowed access to unlinked data sets? In what way?
3. a) If you currently allow access to your data holdings, how are decisions about access adjudicated?

b) Is the adjudication for access to linked data different from unlinked data? In what way?
 4. If you currently allow access to your data holdings, what conditions do you set for accessing the data?

a) Are the data available to individuals to use outside the organization or must individuals access and use the data on the organization's premises?

b) Do you provide technical support to individuals/organizations who have been given permission to access your databases?

c) Do you have any tools, such as data documentation and algorithms, that help data users understand the data and changes over time, and generally make the data more useful to users. For example, a Concept Dictionary is documentation about standardized approaches and definitions.

c) Are there charges to access the database? If so, on what are they based and do they cover the costs of providing access? Are there ways in which these charges can be reduced?
 5. Please describe some of the major difficulties/barriers you face in granting access to your data holdings both within and outside your organization. What resources or supports, and changes would you need to be able to increase access to unlinked and linked data holdings?
 6. What are some of the common complaints individuals and organizations have expressed regarding getting access to your data holdings? What suggestions do you have that would alleviate these difficulties and improve access for individuals or organizations?

Data linkages

This section deals with how data are currently used. Often in conducting population-based health and health services research, researchers would like to link and integrate data from different sources at the regional, provincial, and/or the national level; e.g. linking data from the National Population Health Survey with provincial hospital, physician, or home care utilization data. Differences in data standards, the diversity of access policies and procedures and the need for individual approvals for each data set across jurisdictions are often cited as barriers to such linkages.

1. a) Are data linkages possible with other databases either within your organization, across organizations, or across jurisdictions (regional or provincial)? Please describe.

b) Do you currently have any collaborative arrangements with other jurisdictions in the collection and use of databases?

c) Do you think data linkage is an important issue/area to pursue? Are there databases with which you would like to link but are currently not able to do so?

2. a) What are the major barriers to data linkages with other databases within your organization, across organization, or across jurisdictions? Some barriers may include:
 - lack of standards in data definition and measurement
 - technical problems
 - lack of, or differences in data access policies and procedures
 - differences in privacy and confidentiality rules
 - lack of necessary resources
 - lack of leadership, etc.
- b) What suggestions do you have for overcoming these barriers/disincentives?
3. Are there opportunities for greater data linkage? If so, how can this best be achieved?

Conceptual framework for population-based health and health service research databases, registries, and repositories in Canada

Based on a review of frameworks and taxonomies that have been used to develop an inventory of databases, we are developing a conceptual framework that can serve as a template for the systematic collection of data about Canadian population-based health and health services research databases, registries and repositories. Please find a draft version of our conceptual framework in Appendix 1.

1. From your point of view, does this framework capture the important categories for an inventory of population-based health and health service research databases? Are there any categories and subcategories that should be deleted, added or modified from the ones we have listed?
2. We are interested in testing if the categories and subcategories listed capture all the relevant elements of your data sets. Would you be willing at your convenience to try and classify/catalogue your data holdings in terms of this conceptual framework?

Electronic inventory for population-based databases

Our intention is to use the conceptual framework to drive the development of a prototype of an electronic “meta database,” i.e. a database of databases, to classify existing and potential databases. Ultimately such an electronic inventory would be web-based, would describe various characteristics of existing databases, and could serve as a resource for researchers and other interested individuals/organizations to identify potential research resources.

1. From your perspective, do you believe such an inventory would be a useful resource for data collectors, custodians, and users? Do you anticipate any problems or difficulties for your organization if such a resource existed?
2. a) If you believe that a future electronic inventory would be useful, what essential elements should be built into the prototype that would provide the necessary information about population-based health and health services research databases for various users ?

b) How difficult would it be to provide the information to populate your recommended data elements?

3. What are the key capabilities that such a resource should have? Do you know of any such inventory(ies) that could stand as best practices and what are the features that make it/them useful?
4. Who should have ongoing responsibility for the maintenance, costs and updating of the inventory, and in particular, for ensuring its accuracy and functionality?
5. What do you see as the major potential barriers in the development, maintenance and updating of an electronic inventory?

Overall comments

1. From your perspective, what are the major gaps in Canadian population-based health and health services research data? Where would you suggest CIHR and other health research granting agencies need to place their strategic investments?
2. What are the three or four major barriers in developing and/or maintaining population-based databases so that they can be accessed by users?
3. What are the three or four institutional barriers to developing a coordinated, national approach/inventory for population-based health and health services research databases?
4. From your perspective, what are the priorities for future investments in Canadian population-based health and health services databases in order to support high quality research?

Interview schedule for researchers/users of population-based health and health services data

Background

We are interested in learning about your research needs with respect to population-based health and/or health services data; your experience accessing and using such databases, registries, or repositories; your views on the development of a prototype for an electronic inventory of such databases, repositories and registries; and your comments on future investments in such resources.

Data needs

1. Do your research interests require the analysis of Canadian population-based health and/or health services data? If so, could you describe which databases you typically access?
2. Are there research questions/areas on which you would like to work but are unable to
 - because of the lack of appropriate data sets?
 - because of inability to locate appropriate data sets?Please describe.
3. Have you ever used linked data sets in your research? If not, would you like to use such data?

Data access

1. In your research what are some of the challenges or barriers you face or have faced in using/accessing databases? The list below contains some examples of problems that have been cited in the literature. Could you please expand on your experiences?
 - limited usefulness of data released due to the suppression of variables or cells
 - restricted access
 - onerous access process, i.e. need for multiple approvals, multiple approvals across jurisdictions and/or sectors, approval times
 - need to get consents from multiple data custodians or from individuals whose information is included in the database
 - financial costs
 - data quality
 - non-standardized or lack of clear definitions
 - lack of technical support in using database
 - restrictions on site where data must be accessed
 - restrictions on analysis and/or publication
 - other issues
2. Have you had difficulty, specifically, in accessing linked data sets? What are/were the barriers?
3. Do you have any suggestions for alleviating the problems you have had and improving access for both linked and unlinked data sets?

Data security

1. What is your current understanding about data privacy legislation/regulations/rules and what do you perceive your responsibilities as a researcher and data collector/user to be in this regard? How do you find out about these rules and your responsibilities and changes in them?
2. When collecting or using research data, you become a custodian of those data. Do you perceive your responsibilities as a collector of these data to be different from that as a custodian of these data? If so, in what way?
3. What are the barriers to being current in your knowledge and understanding of data privacy legislation/regulations/rules and your responsibilities as a user of data?
4. When your research is completed, what do you do with your data?

Conceptual framework for population-based health and health service research databases, registries, and repositories in Canada

Based on a review of frameworks and taxonomies that have been used to develop an inventory of databases, we are developing a conceptual framework that can serve as a template for the systematic collection of data about Canadian population-based health and health services research databases, registries and repositories. Please find a draft version of our conceptual framework in Appendix 1.

1. From your point of view, does this framework capture the important categories for an inventory of population-based health and health service research databases? In other words, in attempting to locate data sets for your research, would these categories lead you to the appropriate resources?
4. Are there any categories and subcategories that should be deleted, added or modified from the ones we have listed?

Electronic inventory for population-based databases

Our intention is to use the conceptual framework to drive the development of a prototype of an electronic “meta database,” i.e. a database of databases, to classify existing and potential databases. Ultimately such an electronic inventory would be web-based, would describe various characteristics of existing databases, and could serve as a resource for researchers and other interested individuals/organizations to identify potential research resources.

6. From your perspective, do you believe such an inventory would be a useful resource for data collectors, custodians, and users? Do you anticipate any problems or difficulties if such a resource existed?
7. If you believe that a future electronic inventory would be useful, what essential elements should be built into the prototype that would provide the necessary information about population-based health and health services research databases to you as a user?
8. What are the key capabilities that such a resource should have? Do you know of any such inventory(ies) that could stand as best practices and what are the features that make it/them useful?

9. Who should have ongoing responsibility for the maintenance, costs and updating of the inventory, and in particular, for ensuring its accuracy and functionality?
10. What do you see as the major potential barriers in the development, maintenance and updating of an electronic inventory?

Overall comments

5. From your perspective, what are the major gaps in Canadian population-based health and health services research data? Where would you suggest CIHR and other health research granting agencies need to place their strategic investments?
6. What are the three or four major barriers in developing, maintaining, and accessing population-based health and health services research databases?
7. Are there opportunities for greater data linkage? If so, how can this best be achieved?
8. What are the three or four institutional barriers to developing a coordinated, national approach/inventory for population-based health and health services research databases?
9. From your perspective, what are the priorities for future investments in Canadian population-based health and health services databases?

APPENDIX D: Search strategy for access, privacy and confidentiality issues

The objectives of this phase of the project are:

1. To identify the range of issues and challenges faced by researchers when accessing and using existing population-based health and health services research data;
2. To identify the range of issues and challenges faced by population-based health and health services research data custodians when granting data access to researchers.

References were identified by:

- Searching PubMed;
- Scanning the websites of key stakeholders (e.g. CIHR, CIHI, CMA, research organizations who are data custodians);
- Circulating a request for information to stakeholder groups, the funding organizations and on list serves;
- Collecting resources brought to our attention by the health research and health data communities in Canada.

Using a “snowball” technique, the bibliographies of these references were searched to identify additional articles of relevance.

Searching PubMed

The purpose of this review was to identify the issues involved in using population-based health and health services data for research rather than to present a comprehensive review of this literature or to discover issues with particular data sources or indicators.

There were two main challenges in designing the search strategy. First, there is a multiplicity of terms that can be used to refer to the secondary use of data for research purposes. For example, the phrase “secondary use” is not found in PubMed, whereas the phrase “empirical research” yielded 3,374 articles and did not appear to isolate ones on the topic of interest.⁵⁹

Second, many articles that report on quantitative health or health services research include a discussion of the data used and challenges encountered including issues of data access, comparability and quality. But in most cases, the access or use of the data is not the main topic of the article.

The search strategy was designed to identify articles that deal with common issues in the secondary use of data for health-related research as a main topic. In other words, those articles that address all of the following three components:

- Research
- Data

⁵⁹Searches performed January 14, 2004.

- Issues related to the access or use of data including privacy, confidentiality, consent, disclosure, computer security, data storage or intellectual property.

As a first step, the MeSH (Medical Subject Heading) database in PubMed was used to map search terms to the standardized subject headings used in PubMed. Table 1 lists the terms of interest and the relevant corresponding MeSH terms.

Table 2 depicts the search strategy. The terms listed within each column were combined with OR to capture the variety of terms that PubMed uses to refer to research, data and data access issues respectively. The columns were then in turn combined with AND to identify those articles that cover all three of these concepts. The search was restricted to the MeSH Major Topic field and was limited to Canadian articles by adding the word (not the MeSH term) “Canada” to the entire strategy. No date limits were used. This search strategy identified 230 articles for which we reviewed abstracts where available.

In addition to the above, PubMed was searched for relevant articles by the following authors:

- Charlyn Black
- Tim Caulfield
- Ann Cavoukian
- David Flaherty
- Daniel Friedman
- Chuck Humphrey
- Patricia Kosseim
- William Lowrance
- David Loukedelis
- George Radwanski
- Leslie Roos
- Noralou Roos
- Diana Royce
- Jennifer Stoddart
- Wendy Watkins
- Jack Williams
- Don Willison

Table 1: Terms used in PubMed search

Project concepts/search terms	Relevant MeSH terms
Research	Research Empirical Research Health Services Research Research Support
Data	Data Collection Data Interpretation, Statistical Databases, Genetic Research Design Computer Security Databases, Factual Data Compression Information Storage and Retrieval Database Management Systems Databases Information Systems Confidentiality Questionnaires Nutrition Surveys Empirical Research Ethics Committees, Research Censuses Disease Notification Hospital Records Health Surveys Diet Surveys Dental Records Dental Health Surveys
Personal Health Information	Medical Records
Population-based Data	None
Health Data	Dental Health Surveys Health Surveys Nutrition Surveys
Health Services Data	None
Secondary (use)	None
Privacy	Privacy Confidentiality Genetic Privacy Access to Information Patient Access to Records
Confidentiality	Confidentiality Genetic privacy
Research Ethics Board	Ethics Committees, Research

Project concepts/search terms	Relevant MeSH terms
Consent	Informed consent Consent Forms Third-party consent Disclosure
Collection	Data Collection Health Surveys Surveillance System Health Status Indicators Nutrition Surveys Population Surveillance Health Care Surveys Questionnaires Records Medical Records Medical Record Linkage Medical Records Systems, Computerized Registries Vital Statistics Empirical Research Databases, Factual Databases, Genetic
Access	Access to Information Patient Access to Records Disclosure Mandatory Reporting Computer Security
Data storage	Information Storage and Retrieval Data Compression
Disclosure	Disclosure Mandatory Reporting
(Data) infrastructure	None
Linkage	Data Collection Medical Record Linkage
Capacity development	None
Skills	None
Data quality	Research Design
Accuracy	Population Surveillance
Intellectual property	Intellectual Property Copyright Patents
(Data) completeness	None

Project concepts/search terms	Relevant MeSH terms
(Data) timeliness	None
(Data) stewardship	None
(Data) comparability	None
(Data) disposal	None

Note: Where there are more specific terms under the terms listed above, these specific terms were also included in the search.

Table 2: Search strategy for PubMed search

Relevant MeSH terms for research	Relevant MeSH terms for data	Relevant MeSH terms for data access issues
Research	Data Collection	Privacy
Empirical Research	Data Interpretation, Statistical	Confidentiality
Health Services Research	Research Design	Genetic Privacy
Research Design	Computer Security	Consent Forms
Research Support	Databases, Factual	Third-party Consent
Ethics Committees, Research	Data Compression	Informed Consent
	Information Storage and Retrieval	Disclosure
	Database Management Systems	Ethics Committees, Research
	Databases	Computer Security
	Information Systems	Information Storage and Retrieval
	Confidentiality	Data Collection
	Questionnaires	Intellectual Property
	Empirical Research	
	Ethics Committees, Research	
	Censuses	
	Disease Notification	

Relevant MeSH terms for research	Relevant MeSH terms for data	Relevant MeSH terms for data access issues
	Hospital Records	
	Health Surveys	
	Diet Surveys	
	Dental Records	
	Dental Health Surveys	

Note: Where there are more specific terms under the terms listed above, these specific terms were also included in the search.

Survey of Stakeholder Websites

The following organizations' websites were also scanned for reports related to the use of population-based health and health services data for research purposes:

- Centre for Health Services and Policy Research (www.chspr.ubc.ca)
- Canadian Institute for Health Information (www.cihi.ca)
- Canadian Institutes of Health Research (www.cihr.ca)
- Canadian Medical Association (www.cma.ca)
- Health Canada (www.hc-sc.gc.ca)
- Institute for Clinical Evaluative Sciences (www.ices.on.ca/)
- Interagency Advisory Panel on Research Ethics (www.pre.ethics.gc.ca)
- Manitoba Centre for Health Policy (www.umanitoba.ca/centres/mchp/)
- National Aboriginal Health Organization (www.naho.ca)
- Population Health Research Unit (phru.medicine.dal.ca/)
- Privacy Commissioner of Canada (www.privcom.gc.ca/)
- Social Sciences and Humanities Research Council of Canada (www.sshrc-crsh.gc.ca/)
- Statistics Canada (www.statcan.ca)

APPENDIX E: Literature search strategy and other methods for Chapters 4 and 5

Literature review

A systematic, though not exhaustive, literature search was conducted to identify key published and unpublished literature discussing existing frameworks, taxonomies and projects used to develop, or involved with developing inventories of population-based health and health services research data. Specifically, the search attempted to identify literature in English discussing existing or conceptual databases, registries and repositories, regardless of geographic jurisdiction or date of publication (to date of search, October 2003).

Medline, a health-related database with international coverage from the US National Library of Medicine, was searched for articles and papers to develop and test the search strategy. Subsequently, the search was conducted in Medline, Cochrane Library, Dissertation Abstracts, Embase, PubMed, and World of Science. Key concepts were searched using subject headings and text words appropriate to the individual databases (Table 1).

In addition to searching of conventional databases, grey literature was identified by searching a number of health-related library catalogues, gateways, research organization websites and search engines. These included:

- Canada Institute for Scientific and Technical Information (CISTI) Library Catalogue
- Canadian Institute for Health Information (CIHI)
- CMS Data website, Centers for Medicare and Medicaid Services (US)
- Computer-assisted Survey Methods Program website at the University of California (Berkeley)
- Health Canada
- National Committee on Vital and Health Statistics, Data Council, Department of Health and Human Services (US)
- NLM Gateway (National Library of Medicine, US)
- Data Documentation Initiative, an international organization with its website hosted at the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan, Ann Arbor
- nesstar website, owned and operated by the UK Data Archive and the Norwegian Social Science Data Services
- Research Data Assistance Center (ResDAC) at the University of Minnesota, Minneapolis
- Search engines: Google, Vivisimo

Table 1: Search terms for literature review

Project concepts	Relevant Medical Subject Headings or Text Words
Population-based health or Health services research	MeSH: Population Surveillance Health Services Research Text words: population-based population health health services research health policy research
Data inventories, etc.	MeSH: Databases Databases, Factual Medical Records Medical Record Linkage Medical Records Systems, Computerized Registries Information Systems Management Information Systems Geographic Information Systems Text words: database* databank* dataset* data (base* OR bank* OR set* OR laborator*) registry registries record (linking OR linkage) linked record* repositor* (health OR medical) (record OR records)
Taxonomies or frameworks	MeSH: Models, Theoretical Text words: (theoretical OR theoretical) model* framework* taxonom* categoriz* categoris* classification* classify* classification system*

Discussions with key informants

In addition to the formal literature review, team members made contact with key informants who could both identify additional resources for consideration and provide more in-depth knowledge about relevant resources identified. For example, the team met with Chuck Humphrey, Data Library Coordinator at the University of Alberta, to gain a more in-depth understanding of the Data Documentation Initiative, an international effort to establish a standard for technical documentation describing social science data. A telephone interview was also conducted with Nick Black, Professor of Health Services Research at the London School of Hygiene and Tropical Medicine, who has developed the Directory of Clinical Databases (DoCDat).

We also met, by telephone or in person, with key individuals to enhance our understanding of processes underway in Canada. Michael Wolfson, Assistant Chief Statistician, Analysis and Development Field at Statistics Canada, identified relevant resources and assisted with interpretation of approaches Statistics Canada has undertaken to enhance data availability for the research community. Contacts were made with a number of individuals at Health Canada to identify additional resources and to develop a clearer understanding of some of the data and database inventory activities underway, including Greg Sherman, Director of the Infostructure Development Division (Population and Public Health Branch); Ora Kendall, Chief of the Data Development and Exchange Program (Population and Public Health Branch); Elise Lavigne, Project Manager of the Canadian Public Health Infospace program (Population and Public Health Branch); Alain Vaillancourt, Portal Librarian in the Infostructure Development Division (Population and Public Health Branch); and Bill Bradley, Director of Data Systems and Standards in the Information, Analysis and Connectivity Branch. Interviews were also conducted with key informants at the Canadian Institute for Health Information to provide perspective about CIHI's role in developing database inventories and other tools to support the research community. These included Jennifer Zelmer, Vice-President Research and Analysis; Greg Webster, Director Research and Indicator Development; Louise Ogilvie, Director of Health Services Information; and Steve Slade, Consultant Physician Databases. An interview also took place with Glenda Yeates, CIHI's new President and Chief Executive Officer. Richard Alvarez, former President and CEO of CIHI provided historical perspective about CIHI's evolving interaction with the research community. From his perspective as the new President and CEO of Canada Health Infoway, he provided an understanding about the organization's mandate and future opportunities to build in approaches to support research access in systems that they support. Finally, interviews were also conducted with the Director of CIHR's Institute of Population and Public Health, John Frank, and with the Assistant Director of CIHR's Institute of Health Services and Policy Research, Diane Watson.

Developing an expanded set of resources, critical analysis and synthesis

To ensure that the team identified a relatively comprehensive set of relevant resources, and from this, a nuanced understanding of the significance of various frameworks and activities, considerable effort was invested in an iterative process of identifying, interpreting and classifying resources.

Identification of resources related to conceptual frameworks and taxonomies

For this component of the project, the search outlined above was augmented with discussions with key informants, including members of the Steering Committee, and active exploration of websites identified as potentially relevant. This led to the addition of resources that were relevant, but not captured in the formal search process—e.g. the Wilks report on health information and the associated Template for Health Information, mandates of the CIHR Institutes of Health Services and Policy Research and Population and Public Health, background conceptual

frameworks pertaining to population health, public health, surveillance and epidemiological data systems, and health services research. This led to an expanded set of resources that provided information both about relevant conceptual frameworks for documenting content coverage of data sets, but also to identification of variables to categorize data holdings that would form important components of an inventory of databases.

Identification of resources related to development of electronic inventories

Building from resources identified in the literature review and discussions with key informants, we identified websites and data inventory projects that could serve as examples of best practice for development of electronic inventories. These websites went far beyond a concentration on health and health care, although the few we found that had a health focus were ultimately among the most useful to us. We reviewed websites until additional websites were no longer adding new information.

From the resources identified and the discussions with key informants, the research team identified four specific areas that required critical analysis and synthesis for this project. The first involved a review and synthesis of research-oriented conceptual frameworks in the areas of population health and health services research. These frameworks provided useful input for the development of a section of the data collection tool that could collect information about the content of data holdings, in relation to a combined population health and health services research framework. A second set of resources, gleaned from developmental or ongoing inventory projects, provided information about important domains (i.e. beyond database content) for which an inventory must collect information; these information sources also provided potential wording for development of the data collection tool. A third set of resources provided information about current data and database inventory activities in international, national and provincial settings. A fourth set of resources was used to identify best practices and potential options for development of an electronic inventory, by classifying resources identified into representative categories.

Development of tools for creation of an electronic inventory

Development of a data collection tool

The project team developed a conceptual model of information relevant for inclusion in an inventory of data resources by reviewing existing inventories and literature. This process involved mapping key areas of content and specific items common to inventories, and identifying areas where new areas and/or items were required.

This activity focused on two major domains. The first was a review of existing conceptual frameworks to describe important constructs, relationships and research areas in the combined areas of population and public health. This was an important input for developing a map of the content areas for which an inventory of databases should potentially provide information. It was necessary because while there are separate existing conceptual frameworks for population health and for health services research, there is no widely accepted framework that integrates these two areas. Based on this conceptual framework, a set of items was developed to capture information about the content areas of a range of data sources.

In addition to being able to capture information about the content area covered by data sources, an inventory must provide standard information about a given data set, how it is maintained, the form in which data are collected or available, and how researchers can access the resource. We reviewed existing data inventories, including some non-health inventories, to identify these more generic areas, both in terms of general content and specific items. In addition, we identified areas where new content was required because of the unique aspects of the population

health and health services research domain, with inherent challenges of developing an inventory across such a diversity of data holdings (for example, by developing an item about the derivation of the data source from survey, administrative data or multiple sources).

From these two sets of activities, we developed a data collection tool that could be used to collect information about individual data holdings (i.e. metadata), and that would ultimately provide the basic tool for collection of data about data sets, the activity central to development and maintenance of an inventory.

Development of a framework to screen databases for inclusion in an inventory

The RFP outlined databases to be included in an inventory as comprising administrative databases, registries and surveys. The research team identified a need to develop a tool that would provide additional guidance about which data resources should be included in an inventory. A starting point for development of this tool was an inventory of data resources developed in Alberta, which provided an initial framework for inclusion and exclusion of candidate data sets. This framework was modified by the research team to produce a flowchart that identifies a broad range of data sets relevant for population health and health services research as eligible, but excludes less relevant data sets.

Pilot testing of tools for an electronic inventory

We tested the data collection tool on a sample of 5 databases, representing a range of resources that would be eligible for inclusion in a population health and health services research database inventory. Our intent was to test the ease and utility of the data collection tool and to ensure that it was capable of capturing meaningful information for a range of academic disciplines across a wide array of data types.

Selection of databases for review

Given our interest in testing the conceptual framework across a range of eligible data sources, we identified discrete categories of data sources that would be eligible for inclusion in an inventory and tested five data sets representing a mix of national and provincial/regional sources.⁶⁰

Implementation of data collection tool on databases selected

To test the conceptual framework (now functioning as a data collection tool), we took the following steps:

1. Locate online resources for each of the data sets identified.
2. Fill in as much of the data collection tool as possible using those resources.
3. Identify a contact who could provide additional information as required (e.g. on contents, if not available online) and who would be willing to review the final version of the data collection tool.⁶¹
4. Review results to identify any modifications required for the data collection tool.

⁶⁰Given the time and resources available for this project, we also took into consideration databases about which we already knew something, or for which we were confident we could find a contact for review.

⁶¹We did not identify a resource for review for the Statistics Canada or CIHI data sets, as there was sufficient information available online for these data sets.

APPENDIX F: Data collection tool – framework for review of databases

* DENOTES MANDATORY FIELD

1. GENERAL INFORMATION

* Database Name:

* General description:

* Purpose:

* Sponsor/Collector/Custodian:

Describe the timeframe covered:

General description of reference population and/or geographic area covered:

Data collection methods:

Changes in data over time/data updates:

Outputs (including analysis, reports and publications):

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.):

Approximate number of records in database per year:

When and how often are data collected and how are data made available?

Language:

Funding agency and grant number (if applicable):

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

* Data source: (check all that apply):

Survey

Note: If survey, identify source of information:

Primary respondent report

Other informant report

Direct observation

Clinical records

Administrative records

Registry (e.g. population, disease, profession)

- Vital statistics
- Census
- Other _____

*** Representativeness/population coverage:**

- Total population
- Sample with weights
- Other _____

*** Temporal nature of data:**

- Longitudinal
- Cross-sectional

*** Level of information collected/unit of observation:**

- Individual

Note: If individual, then note type of identifiability:

- Identifiable
- Reversibly anonymized
- Irreversibly anonymized

Note: If individual, then note nature of consent for use:

- Subjects informed individually and written consent processes exist
- Subjects informed individually and opt-out processes exist
- Subjects informed collectively about data collection and use
- Subjects not informed explicitly
- Other _____

- Aggregate

Note: If aggregate, then also note scale (check all that apply):

Individual

- Family
- Workplace
- Municipality
- Province
- National
- Other _____

Note: If aggregate, then also note underlying data structure:

- Individual-level, aggregated
- Aggregate/contextual

*** Potential for data linkage (check highest level possible):**

- Person-specific, longitudinal linkage to other databases is possible
- Aggregate or contextual level linkage (e.g. using three digit postal code, etc.) to other databases is

- possible
- Record level linkage within the database is possible
 - No record linkage is possible, either within the dataset or to other data sets

* **Years covered/available:** _____

3. * DATA CONTENT

Individual/population characteristics:

Health status

- Generic health status
- Health perceptions
- Well-being
- Impairments of body functions (WHO-International Classification of Functioning, Disability and Health)
- Impairments of body structures (WHO- International Classification of Functioning, Disability and Health)
- Activity limitations and participation restriction (WHO- International Classification of Functioning, Disability and Health)
- Health/disease condition(s)
- Mortality
- Summary index

Socioeconomic status

- Educational attainment
- Economic position
- Labour market
- Consumption patterns

Biological factors

- Genetic predispositions
- Immune response
- Cardio-vascular fitness
- Blood chemistry
- Nutritional status

Psycho-social factors

- Personal efficacy
- Personal resources
- Family and friends
- Workplace
- Acculturation
- Coping skills/resilience

Cognitive factors

- Beliefs
- Knowledge
- Attitudes
- Temperament

Behaviours

- Substance (ab)use
- Risk taking
- Physical activity
- Eating
- Compliance
- Sleeping

Exposures

- Dwelling
- School
- Workplace
- Outdoors
- Automobile
- Other transit

Demographics

For individual level:	For population level:
<ul style="list-style-type: none"> <input type="checkbox"/> Age <input type="checkbox"/> Sex <input type="checkbox"/> Family <input type="checkbox"/> Geography/community <input type="checkbox"/> Workplace <input type="checkbox"/> School <input type="checkbox"/> Cultural affiliation 	<ul style="list-style-type: none"> <input type="checkbox"/> Population size <input type="checkbox"/> Age distribution <input type="checkbox"/> Sex ratio <input type="checkbox"/> Socioeconomic characteristics <input type="checkbox"/> Geographic characteristics

*External milieu/factors that influence health***Education**

- Educational attainment
- Funding
- Private schools
- School characteristics
- Community climate

Political

- Civic participation
- Political structure
- Power groups

Environmental

- Air quality
- Water quality
- Food safety
- Physical safety
- Land use
- Environmental hazards
- Natural disasters
- Other

Housing

- Housing stock
- Residential patterns
- Regulation
- Financial issues

Governmental

- Funding
- Policy/legislation
- Services

Socio-cultural

- Social support
- Geographic mobility
- Recreational facilities/green space
- Political
- Volunteer organizations
- Union participation
- Charitable giving
- Protective services

Behavioural

- Tobacco use
- Physical activity
- Diet/obesity
- Alcohol and illicit drug use
- Violence and criminality

Transport

- Safety
- Infrastructure
- Traffic patterns
- Vehicles
- Public transportation
- Funding issues

Employment

- Employment/unemployment rates
- Workforce characteristics
- Area business capacity
- Job access
- Occupational safety
- Job quality
- Job characteristics

Economic

- Income
- Wealth
- Poverty
- Economic development
- Financial services
- Cost of living
- Redistribution
- Fiscal capacity
- Exploitation

Health care, public health and other health interventions

Supply and capacity of the health system

- Human capital
- Facilities
- Equipment
- Training
- Research
- Information systems

Access

- Availability
- Accessibility
- Accommodation
- Affordability
- Acceptability

Use/cost/expenditures

Aspect covered

- Use
- Cost
- Expenditure

Type of care/service

- Public monitoring (e.g. food inspection, surveillance activities)

- Preventive (e.g. screening, immunization, education)
- Primary care
- Diagnostic services
- Specialty care
- Tertiary/quaternary care
- Emergency services
- Home care services
- Mental health care
- Long term care
- Oral health care
- Alternative care
- Palliative care
- Pharmaceuticals
- Ambulance
- Public health
- Rehabilitation
- Hospital care

Setting

- Hospital (includes inpatient, outpatient, day surgery, emergency, clinic)
- Other health care institutions (e.g. practitioner's office, nursing home, hospice)
- Private dwelling
- Community
- Other _____

Provider

- Physician FP
- Physician specialist
- Nurse
- Midwife
- Dentist
- Other allied professional\
- Family member
- Friend
- Other provider _____

Performance

- Equity
- Effectiveness
- Quality
- Safety/adverse events
- Competence

- Continuity
- Appropriateness
- Organization of care
- Case mix adjustment
- Satisfaction
- Waiting times
- Efficiency

Health care system related outcomes

- 30 day mortality
- Readmission rates
- Other _____

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

* **Contact:**

* **Organization housing or maintaining the data source:**

Associated link/URL:

Protocols that govern access to data:

* **Is there a data request process/form for researchers?**

Is a data dictionary available?

Service charges:

Timeline to access data:

Training and support available for researchers:

Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers.

APPENDIX G: Pilot testing the data collection tool

Example 1: Early Development Instrument (EDI)

DATA COLLECTION TOOL: FRAMEWORK FOR REVIEW OF DATABASES

1. GENERAL INFORMATION

General description: The EDI data are gathered on children at the kindergarten level to identify patterns of children's vulnerability based on five domains of interest:

1. Communication skills and general knowledge;
2. Emotional maturity;
3. Language and cognitive development;
4. Physical health and well-being; and
5. Social competence.

Purpose: The EDI project aims to provide school districts and communities with information about their preschool population. The associated mapping project helps:

1. measure readiness to learn in children;
2. assess effectiveness of early childhood interventions; and
3. predict how children will do in elementary school.

Sponsor/Collector/Custodian: The Human Early Learning Partnership at UBC.

Describe the timeframe covered: 2000 onwards.

General description of reference population and/or geographic area covered:

EDI information was first collected in February of 2000 and has expanded to include more districts each year. The number of school districts (out of 59) that participated in the EDI, by school year, is: 3 in 1999/2000, 2 in 2000/2001, 10 in 2001/2002, 43 in 2002/2003, 23 in 2003/2004, 24 in 2004/2005 (anticipated) and 35 in 2005/2006 (anticipated). Every school district participated in at least one round of data collection by the 2003/04 academic year. Some districts have chosen to participate in more than one round of data collection, in some cases to increase their sample size.

Data collection methods:

The EDI-Questionnaire file has every question filled out by all of the teachers for every kindergarten-age child that participated in the survey. The EDI survey consists of 222 variables, organized by the administration and general child information questions and then by three domains: physical health and well being, language and cognitive skills, and social and emotional development. Two further domains (emotional maturity and communication and general knowledge) are covered through derived variables.

Changes in data over time/data updates: Change in sample frame, as described above.

Outputs (including analysis, reports and publications): Early Childhood Development mapping project (<http://www.help.ubc.ca/>).

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.): No information available.

Approximate number of records in database per year: Not available.

When and how often are data collected and how are data made available? Data collected once a year. Made available on request (http://www.edudata.educ.ubc.ca/Data_Pages/EDIsplash.htm)

Language: English.

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

Data source:

Survey

Note: If survey, identify source of information:

Other informant report

Representativeness/population coverage:

Other: Mix of total population and sample of school boards depending on year of data

Temporal nature of data:

Cross-sectional

Level of information collected/unit of observation:

Individual

Note: If individual, then note type of identifiability:

Identifiable

Reversibly anonymized

Note: If individual, then note nature of consent for use:

Other: Parents of subjects informed individually and opt out process exists

Potential for data linkage (check highest level possible):

Person-specific, longitudinal linkage to other databases is possible

Years covered/available: 1999/2000-2003/2004.

3. DATA CONTENT

Individual/population characteristics:

Psycho-social factors

- Personal efficacy
- Coping skills/resilience

Cognitive factors

- Knowledge
- Attitudes
- Temperament

Behaviours

- Physical activity
- Compliance

Demographics

For individual level:
Age Sex Geography/community School Cultural affiliation

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

Contact: Edudata Canada.

Organization housing or maintaining the data source: Edudata Canada.

Associated link/URL: <http://www.edudata.educ.ubc.ca/>.

Protocols that govern access to data: Yes.

Is there a data request process/form for researchers? Yes.

Is a data dictionary available? Yes.

Service charges: Yes.

Timeline to access data: No information available.

Training and support available for researchers: No.

Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers. So far, only internal.

Example 2: CANSIM – Statistics Canada’s Socioeconomic database

DATA COLLECTION TOOL: FRAMEWORK FOR REVIEW OF DATABASES

1. GENERAL INFORMATION

General description: CANSIM includes over 18 million series to help you track trends in virtually every aspect in the lives of Canadians.

- the ability for you to search by subject, keyword, table number or series number.
- increased coverage of socioeconomic data
- and more.

Purpose: CANSIM is an online resource for Canadian socioeconomic statistics on labour, health, income, trade, education, manufacturing, investment and more. It allows you to track trends, analyze market potential or study economic activity with reliable data from the ultimate authority in Canadian statistics.

Sponsor/Collector/Custodian: Statistics Canada.

Describe the timeframe covered: Various; earliest series starts in 1901.

General description of reference population and/or geographic area covered: Canada.

Data collection methods: Data are derived from a wide variety of surveys conducted by Statistics Canada.

Changes in data over time/data updates: Dependent on particular information series of interest; changes are documented within the series.

Outputs (including analysis, reports and publications): None in specific—used in a variety of ways.

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.): No information available.

Approximate number of records in database per year: Whole database contains nearly 18 million numeric time series.

When and how often are data collected and how are data made available? Continuous collection and updating.

Language: English and French.

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

Data source:

Survey

Note: If survey, identify source of information:

Primary respondent report

Other informant report

Representativeness/population coverage:

Total population

Sample with weights

Temporal nature of data:

Longitudinal

Level of information collected/unit of observation:

Aggregate

Note: If aggregate, then also note scale (check all that apply):

Family

Municipality

Province

National

Other: industry, occupation, other

Note: If aggregate, then also note underlying data structure:

Individual-level, aggregated

Aggregate/contextual

Potential for data linkage (check highest level possible):

No record linkage is possible, either within the dataset or to other data sets.

Years covered/available: Time series cover 1901 and onwards.

3. DATA CONTENT

Individual/population characteristics:

Health status

Generic health status

Health perceptions

Well-being

Impairments of body functions (WHO-ICF)

Activity limitations and participation restriction (WHO-ICF)

Health/disease condition(s)

Mortality

Summary index

Socioeconomic status

- Educational attainment
- Economic position
- Labour market
- Consumption patterns

Psycho-social factors

- Family and friends
- Workplace

Behaviours

- Substance (ab)use
- Physical activity
- Eating
- Sleeping

Exposures

- Dwelling
- Automobile

Demographics

For population level:
Population size Age distribution Sex ratio Socioeconomic characteristics Geographic characteristics

External milieu/factors that influence health

Education

- Educational attainment
- Funding
- Private schools

Housing

- Housing stock
- Financial issues

Governmental

- Funding
- Services

Socio-cultural

- Social support
- Geographic mobility

Union participation

Charitable giving

Behavioural

Tobacco use

Physical activity

Diet/obesity

Alcohol and illicit drug use

Violence and criminality

Transport

Vehicles

Public transportation

Funding issues

Employment

Employment/unemployment rates

Workforce characteristics

Economic

Income

Wealth

Poverty

Cost of living

Redistribution

Health care, public health and other health interventions

Supply and capacity of the health system

Human capital

Equipment

Access

Availability

Affordability

Use/cost/expenditures

Aspect covered

Use

Provider

Physician FP

Physician specialist

Dentist

Other allied professional

Performance

Satisfaction

Waiting times

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

Contact: Data library of participating university/college.

Organization housing or maintaining the data source: Statistics Canada.

Associated link/URL: <http://www.statcan.ca/english/ads/cansimII/atoz.htm>.

Protocols that govern access to data: By request, or through university/college data library.

Is there a data request process/form for researchers? No.

Is a data dictionary available? Some detail is available through the Statistics Canada Integrated Metadatabase.

Service charges: Yes, if not affiliated with a university/college, or if using for commercial purposes.

Timeline to access data: The data are online, so access is nearly instantaneous.

Training and support available for researchers: Yes, online through Statistics Canada (tutorial).

Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers. Yes, internal and external.

Example 3: Canadian Community Health Survey (CCHS)

DATA COLLECTION TOOL: FRAMEWORK FOR REVIEW OF DATABASES

1. GENERAL INFORMATION

General description: The Canadian Community Health Survey (CCHS), is being conducted by Statistics Canada to provide regular and timely cross-sectional estimates of health determinants, health status and health system utilization for 136 health regions across the country.

Funding for the CCHS was provided under the *Health Information Roadmap* initiative, a plan to modernize and standardize health information across the country. The Canadian Institute for Health Information (CIHI) received funds for the *Roadmap* from Health Canada, and Statistics Canada has joined as a partner in supporting a series of projects.

Purpose: The Canadian Community Health Survey (CCHS) – Cycle 1.1, was conducted by Statistics Canada to provide cross-sectional estimates of health determinants, health status and health system utilization for 133 health regions across Canada, plus the territories.

Sponsor/Collector/Custodian: Statistics Canada.

Describe the timeframe covered: 2000/01 onwards, by cycle.

General description of reference population and/or geographic area covered: Canada, with reporting possible down to sub-provincial level.

Data collection methods: The CCHS began collection in September 2000. Each two-year collection cycle is comprised of two distinct surveys: a health region-level survey in the first year with a total sample of 130,000, and a provincial-level survey in the second year with a total sample of 30,000. Sample sizes in any particular month or year may increase due to provincial or health region-level sample buy-ins. Both computer-assisted personal and telephone interviews are used.

The target population of the CCHS includes household residents in all provinces and territories, with the principal exclusion of populations on Indian Reserves, Canadian Forces Bases, and some remote areas. There will be one randomly selected respondent per household, although planned oversampling of youths will result in a second member of certain households being interviewed. For the first collection cycle only those 12 years of age and over are eligible for selection, although it is expected that in future cycles child-specific content will be included.

Changes in data over time/data updates: Data collected will vary to some degree by cycle.

Outputs (including analysis, reports and publications): It is expected that all CCHS products from a particular cycle will be released over the 12 months following completion of the cycle's last interview. A CCHS microdata file will be produced and shared with the provinces, territories and Health Canada under

a data sharing agreement. In addition, a public use file (PUMF) will be produced and released on compact disc. Access to CCHS microdata can also be obtained by using Statistics Canada's custom tabulation and remote access services. Results of each survey cycle will be disseminated in the form of an overview report, quarterly CCHS articles on topics or sub-populations of interest, articles in *Health Reports* and a series of 136 health region profiles available on the Statistics Canada website. Finally, workshops are planned to assist users in maximizing their use of the CCHS.

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.): No information available.

Approximate number of records in database per year: No information available.

When and how often are data collected and how are data made available? Data are collected every year on a rotating survey cycle. Data are made available through yearly data releases.

Language: English and French.

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

Data source:

Survey

Note: If survey, identify source of information:

Primary respondent report

Other informant report

Representativeness/population coverage:

Sample with weights

Temporal nature of data:

Cross-sectional

Level of information collected/unit of observation:

Individual

Note: If individual, then note type of identifiability:

Reversibly anonymized

Note: If individual, then note nature of consent for use:

Subjects informed individually and written consent processes exist

Potential for data linkage (check highest level possible):

Person-specific, longitudinal linkage to other databases is possible.

Years covered/available: 2000/01 onwards.

3. DATA CONTENT

Individual/population characteristics:

Health status

- Generic health status
- Health perceptions
- Well-being
- Impairments of body functions (WHO-ICF)
- Activity limitations and participation restriction (WHO-ICF)
- Health/disease condition(s)
- Summary index

Socioeconomic status

- Educational attainment
- Economic position
- Labour market

Psycho-social factors

- Personal resources
- Family and friends
- Coping skills/resilience

Behaviours

- Substance (ab)use
- Risk taking
- Physical activity
- Eating
- Compliance
- Sleeping

Demographics

For individual level:
Age Sex Family Geography/community Cultural affiliation

External milieu/factors that influence health

Socio-cultural

- Social support

Behavioural

- Tobacco use
- Physical activity

Diet/obesity
Alcohol and illicit drug use

Employment

Employment/unemployment rates
Job quality
Job characteristics

Economic

Income

Health care, public health and other health interventions

Access

Availability
Accessibility
Accommodation
Affordability
Acceptability

Use/cost/expenditures

Aspect covered

Use

Setting

Hospital (includes inpatient, outpatient, day surgery, emergency, clinic)

Other health care institutions (e.g. practitioner's office, nursing home, hospice)

Private dwelling

Community

Provider

Physician FP

Physician specialist

Nurse

Dentist

Other allied professional

Performance

Quality

Satisfaction

Waiting times

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

Contact: Data library at university/college participating in the Data Liberation Initiative (for Public Use Masterfile), or Statistics Canada Research Data Centres (for Microdata files).

Organization housing or maintaining the data source: Statistics Canada.

Associated link/URL: <http://www.statcan.ca/english/concepts/health/cchsinfo.htm>.

Protocols that govern access to data: Yes.

Is there a data request process/form for researchers? Yes, for access to data through Research Data Centres.

Is a data dictionary available? Yes.

Service charges: No.

Timeline to access data: Within 8 weeks of application (<http://www.statcan.ca/english/rdc/apply.htm#Proposal>).

Training and support available for researchers: No.

Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers. Yes, internal and external.

Example 4: Discharge Abstract Database – CIHI

DATA COLLECTION TOOL: FRAMEWORK FOR REVIEW OF DATABASES

1. GENERAL INFORMATION

General description:

The Discharge Abstract Database (DAD) contains data on hospital discharges across Canada.

Purpose:

The purpose of DAD is to:

- support CIHI's mandate;
- collect, process and analyse summaries of hospital discharges and day surgeries;
- support management decision making at the hospital, regional and provincial/territorial levels;
- facilitate provincial and national comparative reporting;
- support the development and use of analytical tools, such as case grouping methods, length of stay analysis and resource utilization analysis;
- support related approved analysis and research by others.

Sponsor/Collector/Custodian:

CIHI receives data directly from participating hospitals. These include all hospitals in every province and territory, except Quebec. As of fiscal 2004-05, 100% of Manitoba hospitals will be participating.

Describe the timeframe covered.

- Most recent year: 2002/03
- Next release: 2003/04 (December 2004)
- Historical series: 1979/80 – 2002/03

General description of reference population and/or geographic area covered:

This database contains demographic, administrative and clinical data for hospital discharges (inpatient acute, chronic, rehabilitation) and day surgeries. All admissions from all provinces and territories are included.

Data collection methods:

CIHI receives data directly from participating hospitals. These data are abstracted from chart records at participating hospitals according to a protocol maintained by CIHI.

Changes in data over time/data updates: No information available.

Outputs (including analysis, reports and publications):

All clients who submit data to the Discharge Abstract Database may take advantage of CIHI's eCHAP (electronic Comparison of Hospital Activity Program) and electronic Hospital Specific Reports (eHSR). These products are offered at no cost to clients. There is currently no restriction on the number of users at each facility who may access these products although CIHI does reserve the right to do so in the future. Registration is required to access these products and takes less than 5 minutes.

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.):

CIHI strives to ensure that the quality of the information in their data holdings is suited to its intended uses, and that data users are provided good information about data quality.

Approximate number of records in database per year: Not available.

When and how often are data collected and how are data made available?

Data are collected in an ongoing way through reports from hospitals. Databases are compiled on an annual basis using a fiscal year format.

Language: English and some French. Note data dictionary is only available in English.

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

Data source:

Administrative records.

Representativeness/population coverage:

Total population.

Temporal nature of data:

Longitudinal.

Level of information collected/unit of observation:

Individual

Note: If individual, then note type of identifiability:

Reversibly anonymized

Note: If individual, then note nature of consent for use:

Subjects not informed explicitly

Potential for data linkage (check highest level possible):

Person-specific, longitudinal linkage to other databases is possible

Aggregate or contextual level linkage (e.g. using three digit postal code, etc.) to other databases is possible

Record level linkage within the database is possible

No record linkage is possible, either within the dataset or to other data sets

Years covered/available: 1979/80 – 2002/03.

3. DATA CONTENT

Individual/population characteristics:

Health status

Health/disease condition(s)

Mortality

Behaviours

Substance (ab)use

Exposures

Dwelling

School

Workplace

Outdoors

Automobile

Other transit

Demographics

Age

Sex

Geography/community

Cultural affiliation

Health care, public health and other health interventions

Use/cost/expenditures

Aspect covered

Use

Cost

Type of care/service

Hospital care

Setting

Hospital (includes inpatient, outpatient, day surgery, emergency, clinic)

Provider

Physician FP

Physician specialist

Other provider: PT/OT

Health care system related outcomes

Readmission rates

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

Contact: dad@cihi.ca

Organization housing or maintaining the data source: CIHI.

Associated link/URL: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=services_dad_e.

Protocols that govern access to data: Data disclosure is determined by CIHI's privacy principles and policies.

Is there a data request process/form for researchers? Yes. Details for researchers are available at http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=reqdata_snap_e. Details for graduate students are available at http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=reqdata_gsdap_e. The application form for raw data is available at http://secure.cihi.ca/cihiweb/en/downloads/reqdata_e_reqform.pdf.

Is a data dictionary available? Only a variables list is available on the CIHI website.

Service charges: CIHI responds to custom data requests from researchers and others on a cost-recovery basis. Pricing includes a basic administration fee plus production time. CIHI provides health data to qualifying graduate students at no cost.

Timeline to access data: Information not available.

Training and support available for researchers: Through standard CIHI training in related areas, e.g. ICD coding systems.

Have other researchers used the data resource? Note if only used by researchers internal to organization, or if researchers external to organization have used it. Yes, both internal and external.

Example 5: Ontario Cancer Registry

DATA COLLECTION TOOL: FRAMEWORK FOR REVIEW OF DATABASES

1. GENERAL INFORMATION

General description: The Ontario Cancer Registry (OCR) is a computerized database of information on all Ontario residents who have been newly diagnosed with cancer or who have died of cancer. All new cases of cancer are registered, except non-melanoma skin cancer.

Purpose: The Ontario Cancer Registry is used for four main purposes:

- 1) **Research.** The OCR is an invaluable resource for conducting epidemiological studies. It can also be used to help evaluate the efficacy of screening programs. Recently, use of the OCR has expanded to include health care utilization studies. The availability of data regarding utilization of hospitals and cancer clinics by cancer patients as well as details regarding treatment (e.g. surgical procedures) provides researchers with a useful tool in performing these studies.
- 2) **Projecting the future cancer burden.** Cancer Care Ontario (CCO) produces projections on the number of new cancer cases expected in future years. These data are used by CCO and the Ministry of Health to forecast radiotherapy and other patient treatment requirements.
- 3) **Providing cancer data to other agencies involved in cancer surveillance.** The OCR regularly contributes incidence data to the Canadian Cancer Registry based at Statistics Canada, the North American Association of Central Cancer Registries, and to the International Agency for Research on Cancer. The Public Health Branch of the Ministry of Health also receives data from the OCR for its community health information system.
- 4) **Dissemination of descriptive statistics.** Statistics of cancer incidence and mortality are disseminated through special publications and online reports.

Sponsor/Collector/Custodian: Cancer Care Ontario.

Describe the timeframe covered: 1964 forward.

General description of reference population and/or geographic area covered: Ontario residents who have been diagnosed with or died from cancer.

Data collection methods:

The process of cancer registration in Ontario is passive, relying almost completely on records collected for other purposes. Close to 400,000 records are submitted to the OCR each year. Since 1979, the OCR has relied on the same four major data sources:

- hospital discharge and day surgery summaries which include a diagnosis of cancer
- pathology reports with any mention of cancer
- records of patients referred to CCO's eight regional cancer centres or the Princess Margaret Hospital, the specialized institutions treating cancer patients in Ontario

- death certificates, with cancer recorded as the underlying cause of death

All records except pathology reports are coded at the source and provided to the OCR in electronic form. The majority of the pathology reports are electronically transmitted from hospital and private pathology laboratories to the OCR and the diagnosis is then coded by CCO staff. There are still some labs that send paper copies of pathology reports which are then coded and key-entered by OCR staff into the registry.

Changes in data over time/data updates: Regularly scheduled updates from most sources.

Outputs (including analysis, reports and publications):

Examples of descriptive statistics:

- **Cancer Incidence and Mortality in Ontario 1964-1996.** This report describes cancer incidence and mortality trends for lung, breast, prostate and colorectal cancer.
- **Colorectal Cancer in Ontario 1971-1996.** This report, presented at the November 1998 Colorectal Cancer Screening Workshop, describes colorectal cancer incidence, mortality and survival in Ontario.
- **Cervical Cancer in Ontario 1971-1996.** This report describes cervical cancer incidence, mortality and survival in Ontario.
- **Breast Cancer in Ontario 1971-1996.** This report describes breast cancer incidence, mortality and survival in Ontario.
- **Tobacco or Health in Ontario.** This report presents information on incidence, mortality and survival for tobacco-related cancers and on general mortality in Ontario.
- **Ontario incidence data for 1999**
- **Ontario incidence data for 2000**
- **Ontario mortality data for 1999**
- **Ontario mortality data for 2000**

Data quality issues (e.g. general assessment of quality of data, completeness of data, processes in place to detect and correct errors, etc.): Limited control over quality of data received from sources. Changes at data source can affect accuracy, completeness and timeliness of registration of cases.

Approximate number of records in database per year: ~54,000.

When and how often are data collected and how are data made available? Varies by source from daily to quarterly.

Language: English.

2. ATTRIBUTES OF DATA CONTAINED IN DATABASE

Data source:

Clinical records

Vital statistics

Other: Various data sources linked to identify a patient and build the registry

Representativeness/population coverage:

Total population

Temporal nature of data:

Longitudinal

Level of information collected/unit of observation:

Individual

Note: If individual, then note type of identifiability:

Identifiable

Note: If individual, then note nature of consent for use:

Other: Legislative authority to create the registry

Potential for data linkage (check highest level possible):

Person-specific, longitudinal linkage to other databases is possible.

Years covered / available: 1964 – 2002 (1980 onwards is most complete in terms of data elements especially region or county of residence).

3. DATA CONTENT

Individual/population characteristics:

Health status

Health/disease condition(s)

Demographics

For individual level:
Age
Sex
Family
Geography/community

Health care, public health and other health interventions

Health care system related outcomes

Other: Overall mortality

4. DATA AVAILABILITY/ACCESS FOR RESEARCH

Contact: Carole Herbert, 416.971.9800.

Organization housing or maintaining the data source: Cancer Care Ontario.

Associated link/URL: http://www.cancercare.on.ca/research_cancerRegistry.htm.

Protocols that govern access to data: Yes.

Is there a data request process/form for researchers? Yes.

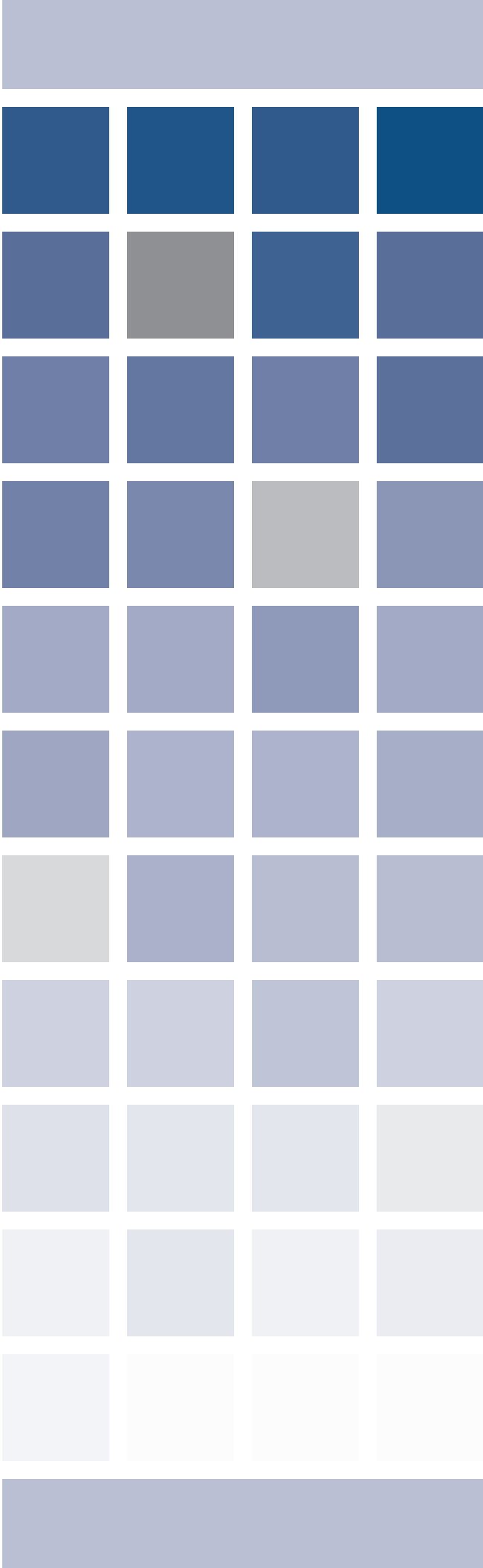
Is a data dictionary available? Yes.

Service charges: Yes.

Timeline to access data: Depends on complexity and timeline for ethics approval (if necessary).

Training and support available for researchers: No.

Have other researchers used the data resource? Note if used only by researchers internal to the organization, or if used by external researchers. Yes, both internal and external.



Centre for
HEALTH SERVICES AND POLICY RESEARCH

Canadian
Policy Research
Networks



Réseaux canadiens
de recherche en
politiques publiques