

BC Centre For Data Innovation

Final Report Of The Working Group

November 2014



About This Report

This report was produced by the BC Centre for Data Innovation Working Group. This group was comprised of members from the BC Ministry of Technology, Innovation and Citizens' Services and the BC Ministry of Health. Deloitte was engaged to facilitate the activities of the Working Group, summarize the outputs of the Working Group in this report and provide subject matter input as requested. The recommendations in this report represent the collective views of the members of the Working Group and are provided for consideration by the Project Sponsors.

Contents

Executive Summary	1
Introduction	3
Opportunity Summary.....	6
Stakeholder Interests	8
Relevant Trends	10
Conceptual Model.....	22
Costs and Funding Options.....	25
Recommendations	28
Appendix A - Overview of Consultation Initiatives	30
Appendix B - Jurisdictional Summaries.....	31
Appendix C - Phase 1 Activities Supporting the Creation of the BC CDI.....	43

Executive Summary

Introduction

Governments generate and collect a vast amount of information associated with the delivery of services to citizens. This information spans multiple domains (health, environment, social services, education, infrastructure, etc.) and, in many cases it has been collected for years, if not decades. The pace at which data is created within each of these sectors is increasing, as is its complexity and variability. This information, often provided by citizens, represents one of the most important assets within government's control. It has the potential to enable advanced research and analysis and to reveal new insights that could provide solutions to some of the greatest challenges facing citizens and governments today. It could also support innovation and economic growth across multiple sectors.

The BC Government recognizes the potential to better utilize public sector information. In December 2013 the Honourable Minister for Technology, Innovation and Citizens' Services, Andrew Wilkinson, and the Honourable Minister for Health, Terry Lake, announced the establishment of a Working Group to evaluate the potential to create a secure centre in BC that would enable more effective use of public sector information for research and innovation purposes. The objective of the Working Group was to identify and evaluate options to support acquisition and linkage of, and access to, data from multiple sectors in a secure manner. The Working Group was asked to initially focus on the health sector, but to ensure that any recommendations be extensible to broader, cross-sector data needs.

The Working Group included representation from the Ministry of Technology, Innovation and Citizens' Services and the Ministry of Health, and was tasked with the following:

- Review the current state of research in BC and identify gaps or needs that are not being met
- Review models used around the world to make public sector data available for research purposes
- Make recommendations on a model that would best address identified gaps and needs while ensuring that:
 - Privacy and security are not compromised in any way
 - All health data is thoroughly de-identified before being made available
 - All data is kept within a secure data centre in BC

Vision For British Columbia

In reviewing similar initiatives in other jurisdictions (in Canada and globally), there was a clear trend toward combining growing public sector data assets with emerging technologies, research techniques and analytical tools to unlock significant value for the benefit of citizens. Whether used for research purposes, to support government policy development and decision making or to enable innovation and economic development, the potential benefits range from improved policy and service delivery outcomes to greater efficiency and effectiveness of government services and broader economic growth.

Based on this, the Working Group developed a proposed vision for the British Columbia Centre for Data Innovation (BC CDI). It highlights the aspirational goals of the centre and recognizes the need to ensure that such a centre works for the benefit of the citizens of British Columbia and does so in a secure and privacy-protecting manner. The proposed vision for the centre is:

A centre of excellence for data that will foster innovation and collaboration between citizens, researchers, the private sector and government. The Centre will enhance the well-being of British Columbians through the responsible use of government data and services grounded upon leading practices in security & privacy.

To enable this vision, the Working Group created a "conceptual model" to illustrate the capabilities and components of such a centre. This model also includes the stakeholders who would interact with the centre as both providers and users of data.

Costing And Funding Options

The conceptual model provides a basis for preliminary, high-level costing of a facility to enable the BC CDI vision. At this early stage, such an exercise can only provide an order of magnitude estimate of such costs. The report outlines estimated capital costs for establishing a net new facility (\$7.0MM-\$10.5MM) and ongoing operating costs (\$4.0MM-\$6.0MM). It also describes the estimated range of costs for onboarding data from new providers and sectors. These costs depend to a great degree on the scale and complexity of the data to be incorporated into the facility, and could range from approximately \$500,000-\$5MM.

The Working Group also explored potential funding sources. These included grants (from provincial and federal agencies and from research funding bodies), contributions from private sector partners, fees paid by participating sectors (for the centre to provide services on their behalf) and user fees.

Recommendations

Recognizing that this is an ambitious, multi-year vision, the Working Group identified immediate actions that could be taken to enhance access to health data for research purposes while building the foundation for the long-term vision. To accomplish this, a series of recommendations was developed to support short-term enhancements while also enabling a concurrent, phased approach to further refining the BC CDI strategy and implementation plan. These recommendations are listed below, with additional detail provided in the body of the report:

1. Approve the vision and value proposition for the BC CDI and initiate a multi-year, phased implementation approach to realize this vision
2. Establish an Interim Governance Committee to oversee the ongoing development of the BC CDI.
3. Leverage existing capabilities and technology to build the foundation for the BC CDI, starting with the Ministry of Health.
4. Broaden the scope of consultation and inquiry to include additional sectors beyond health and additional stakeholders/data consumers beyond academic researchers.
5. Build robust security and privacy protection into each phase of the BC CDI implementation.
6. Streamline the Access Request and Approval Process.
7. Explore short-term opportunity to leverage available government funding for Phase 1 activities while developing a long-term, sustainable funding and revenue model.
8. Evaluate opportunities to leverage emerging technologies and advanced analytics capabilities to maximize the value of public sector data.

Introduction

Project Background and Objectives

On December 5th, 2013, the Minister of Technology, Innovation and Citizens' Services (the Hon. Andrew Wilkinson) and the Minister of Health (the Hon. Terry Lake) announced the establishment of a joint working group to evaluate the potential for a secure data centre in BC. The Working Group was asked to identify and evaluate options to create a BC "Centre for Data Innovation" (BC CDI) that would enable the acquisition, storage and linkage of public sector data in a secure manner. Beginning with the health sector, the Working Group was asked to assess the opportunities and challenges associated with securing, linking and querying data, and to make recommendations that could be expanded to incorporate participants and data from other sectors across government.

Specifically, the BC CDI Working Group was tasked to:

- Review the current state of research in BC and identify gaps or needs that are not being met
- Review models used around the world to make public sector data available for research purposes
- Make recommendations on a model that would best address identified gaps and needs while ensuring that:
 - Privacy and security are not compromised in any way
 - All health data is thoroughly de-identified before being made available
 - All data is kept within a secure data centre in BC

Project Stakeholders

The Working Group included representation from the Ministry of Technology, Innovation and Citizens' Services (MTICS) and the Ministry of Health (MOH, see Figure 1 above). The project was co-sponsored by John Jacobson (Deputy Minister, MTICS) and Dr. Stephen Brown (Deputy Minister, MOH). Strategic direction and oversight were provided by a Steering Committee, which included Bette-Jo Hughes (Associate Deputy Minister, MTICS and BC Government CIO), Lindsay Kislock (Assistant Deputy Minister, Health Sector Information Management/Information Technology) and Heather Davidson (Assistant Deputy Minister, Health Sector Planning and Innovation Division). The Steering Committee, with the support of the Working Group, reported regularly to the Project Sponsors on progress, issues and key decisions for resolution.

The Working Group included multiple representatives from both Ministries as outlined in Figure 1. Representation from MTICS included Charmaine Lowe (Executive Director, Information Stewardship and Policy Branch), Gary Perkins (Executive Director and BC Government Chief Information Security Officer) and Colleen Rice (Senior Policy Advisor). Representation from MOH included Kelly Moran (Executive Director, Information Management and Knowledge Services), Deb McGinnis (Executive Director, Health Information Privacy, Security & Legislation), Victoria Schuckel (Director, Research, Library Services and Knowledge Translation) and Shirley Wong (Director, Information Management and Knowledge Services). Figure 1 highlights the key responsibilities of each group.

Deloitte was engaged to support and facilitate the Working Group's activities. This included providing project management support, conducting research for the jurisdictional scan, providing input on an as-requested basis for specific topics, including technology, security and governance and summarizing the work and recommendations of the Working Group in this report.

Figure 1 - Project Team Structure And Responsibilities

Project Team Members	Responsibilities
Project Sponsors <ul style="list-style-type: none"> • John Jacobson • Dr. Stephen Brown 	<ul style="list-style-type: none"> • Provide strategic direction to the project • Direct the work of the project and review/approve any changes in scope • Make decisions based on recommendations • Provide updates to Ministers as required
Project Steering committee <ul style="list-style-type: none"> • Bette-Jo Hughes • Lindsay Kislock • Heather Davidson 	<ul style="list-style-type: none"> • Provide strategic guidance and oversight to Working Group and ensure it has necessary support • Manage Working Group resources and identify where additional resources are required • Approve budgets as required to support Working Group • Provide updates to Project Sponsors as required
Working Group <ul style="list-style-type: none"> • Charmaine Lowe • Gary Perkins • Colleen Rice • Kelly Moran • Deb McGinnis • Shirley Wong • Victoria Schuckel • Deloitte 	<ul style="list-style-type: none"> • Establish approach and ensure objectives are met within agreed timeline and budget • Coordinate all work streams and support integration and alignment • Report on progress to Steering Committee • Produce final project report for Project Sponsors
Work Streams	<ul style="list-style-type: none"> • Define key requirements for each work stream • Analyse options and propose recommendations • Prepare and validate outputs • Engage key stakeholders as required

Approach

The approach the Working Group took to evaluate the potential for a cross-sector data centre in British Columbia was an incremental one, and was strongly influenced by the following guiding principles:

- Propose a vision for a cross-government data centre, but start with the health sector, creating a foundation which can be expanded to other sectors over time;
- Leverage the significant stakeholder consultation which has already occurred in the health sector to identify key stakeholder interests and concerns;
- Assess the potential to build on existing infrastructure, processes and capabilities in BC where possible; and,
- Identify opportunities to leverage existing funding sources where possible.

To complete its work, the BC CDI Working Group adopted a five-phase approach:

Phase 1 - Project initiation

The purpose of this phase was to confirm the scope and objectives of the project, as well as the approach, activities, milestones and project governance model.

Phase 2 - Jurisdictional scan

In order to inform the development of the vision and conceptual model for the BC CDI, a high-level scan of similar facilities in other jurisdictions was undertaken. The purpose of this scan was to assess how other jurisdictions had approached the development and operation of these facilities and to identify examples of good practice and lessons learned across a number of dimensions. These dimensions included security and privacy, governance, data access, services and capabilities and data management.

Phase 3 - BC CDI Vision, Value Proposition and Conceptual Model

To guide the development of a conceptual model for the centre and to define the desired future state, the Working Group developed a compelling vision for the BC CDI. This vision, combined with identified stakeholder interests, led to the development of a number of expected high-level benefits for citizens and other key stakeholders.

The vision and value proposition for the BC CDI provided a framework within which a conceptual model was developed. The purpose of the model is to define the desired future state in terms of data providers and sources, data management, reporting and analytical functionality, services and proposed end users. This model not only enabled the Working Group to confirm the preliminary scope and scale of the BC CDI, but also provided a model against which to compare options for the recommendations. The development of the model also included high-level estimates of potential costs for such a centre, based on a series of implementation assumptions and on a comparison to costs incurred in other jurisdictions for similar facilities.

Phase 4 - Define Health Sector Data Vision

To support the initial focus on health sector data, the Working Group developed a supporting vision for the health sector to define short-term and long-term objectives for enhanced integration and utilization of data across the sector. The short-term objectives informed the Working Group's recommendations regarding proposed next steps.

Phase 5 - Develop recommendations for next steps

Based on the above work, the Working Group developed a series of recommendations aimed at creating an initial foundation for a BC CDI, which could be expanded, over time, to fully realize the broader goals of the overall vision and value proposition.

Report Overview

This report provides an overview of the work completed by the Working Group and includes the following sections:

Introduction: outlines the background, context and objectives for the project, as well as the project approach and stakeholders.

Opportunity Summary: defines a vision for the BC CDI and describes the high-level benefits of the centre for key stakeholder groups.

Stakeholder Interests: provides a summary of stakeholder interests with respect to the BC CDI based on consultation work conducted as part of the "Support for Patient Oriented Research" business plan process, the "BC Health Research Strategy" process and consultation workshops hosted by the Office of the Information and Privacy Commissioner in 2012 and 2013.

Relevant Trends: provides a summary of relevant trends identified in the scan of similar facilities in other jurisdictions.

Conceptual Model: defines a model that addresses the high-level objectives for the data centre and enables an analysis of options for each component of the solution.

Costs and Funding Options: provides the potential costs and funding sources associated with the creation of the BC CDI. Potential costs are based on costs for similar facilities in other jurisdictions and high-level estimates for the implementation of the conceptual model.

Recommendations: lists the Working Group's recommendations and proposed next steps.

Opportunity Summary

Public sector data holds significant potential for improved transparency and citizen engagement, more efficient and effective service delivery, improved decision making and policy development, and increased innovation and economic growth across a number of sectors. The Province has already realized some of these benefits through its Open Data initiative (<http://data.gov.bc.ca>), which, to date, has published over 3,200 high-value data sets. While the BC Government will continue to make more data open and available for re-use and adaptation on its Open Data site, it is also interested in looking beyond Open Data at other opportunities for unlocking the value of public sector information.

A provincial centre that provides secure access to linkable, anonymized data from across government could be that key, unlocking valuable data in support of research, evidence-based policy and decision-making and product and service development and innovation. This opportunity has the potential to not only address current challenges associated with access to, and linkage of, data, but to realize unprecedented benefits for British Columbians, some of which are outlined below.

While the potential benefits of such a centre are acknowledged, it is also recognized that the single most critical success factor is that the centre be secure and respect the privacy rights of citizens. There are numerous examples of facilities in Canada and internationally where these challenges have been addressed. Further, coordinating and enabling access through one organization and environment can support enhanced privacy protection and security, when compared to a more distributed approach.

The purpose of this section is to outline the vision for the BC CDI, and to highlight the rationale for pursuing this vision by describing the value proposition of such a centre for citizens and other key stakeholders.

Vision

The Province's vision for the BC CDI is set out below:

A centre of excellence for data that will foster innovation and collaboration between citizens, researchers, the private sector and government. The Centre will enhance the well-being of British Columbians through the responsible use of government data and services grounded upon leading practices in security & privacy.

This vision includes several key themes that highlight the aspirations for the centre as well as the requirements it must fulfill. These themes are described in further detail below:

Wellbeing of British Columbians: There is a fundamental recognition that the data collected from and about citizens belongs to citizens, and that all uses should first and foremost drive benefits for them. It is also recognized that there is an expectation that governments use the data at their disposal to improve the well-being of citizens. Whether these benefits are associated with more effective delivery of public services, better engagement with citizens, improved health or environmental outcomes, cost savings, economic growth or other beneficial uses, the goal of any data use should be to enhance the wellbeing of British Columbians.

Innovation and collaboration: The intent for the centre is to create a hub of interaction, exploration, analysis and innovation. It is targeted at a wide range of users to enable enhanced engagement and collaboration between these groups on key questions of interest to each.




Responsible use: Recognizing that the intent of the centre is to create improved outcomes and benefits for British Columbians, data will only be used as permitted under law and where such use provides benefits to the public.

Leading Practices: Requirements for the protection of privacy are clearly outlined in legislation and government policy. Further, security requirements for such environments are described in policy and internationally accepted standards. It is envisioned that the centre will be designed to not only meet these requirements, but to exceed them wherever possible using leading technologies and practices for data management, governance, and decision making, as well as for the protection of privacy and security.

Value Proposition

The rationale for the BC CDI is based on the value proposition it presents to key stakeholder groups in the province. Figure 2, below, provides a qualitative summary of the anticipated benefits associated with the BC CDI by stakeholder group¹

Figure 2 - Summary Of Anticipated Benefits Associated With The BC CDI, By Stakeholder Group

			
Citizens	Researchers	Private Sector	Government
<ul style="list-style-type: none"> • Enhanced transparency and trust • Personalized services and products • More employment opportunities associated with economic growth • Improved outcomes and efficiency of government services (healthcare, social services, environment etc.) • Improved security and privacy of citizen data. 	<ul style="list-style-type: none"> • Access to one of the most robust, linked data resources in Canada • Predictable, timely turnaround on access requests • Increased productivity arising from enabling services and reduced transaction costs for data requests • Improved access to research funding. 	<ul style="list-style-type: none"> • Better understanding of needs of citizens & government • Ability to bring new/enhanced data products & services to market • Enhanced innovation and productivity through the 'cluster effect' • Improved global competitiveness • Improved competitiveness for funding 	<ul style="list-style-type: none"> • Improved understanding of citizen needs • Improved policy development & decision making • Time and resource savings due to more efficient service delivery • Improved data quality • Enhanced innovation through partnerships with citizens and the private sector (co-creation)

¹It should be noted that the reference to "Government" in this figure includes the broader public sector within the province (e.g., school districts, health authorities and other public sector entities).

Stakeholder Interests

As noted in the Introduction of this report, the approach the Working Group took in conducting its review was to start with the health sector and to leverage, and build upon, the significant feedback received through recent stakeholder consultation within the sector. Rather than initiating another round of consultation with the same stakeholders, the outputs of the following three initiatives were used to identify stakeholder interests with respect to data, services and functionality:

- The BC Health Research Strategy (BCHRS), facilitated by the Michael Smith Foundation for Health Research;
- The Support for Patient Oriented Research (SPOR) business plan consultation process (also facilitated by the Michael Smith Foundation for Health Research); and,
- Two stakeholder meetings convened by the Office of the Information and Privacy Commissioner (OIPC): the Health Research Roundtable and Health Data Research Forum on June 25th, 2012 and December 9th, 2013, respectively.

A brief overview of each of these initiatives is provided in Appendix A of this report.

A summary of the key themes of stakeholder interests arising out of these initiatives is set out below. While the context of the consultations was specific to the health sector, many of the themes and interests that surfaced have cross-government applicability and likely overlap with other sectors. Still, moving forward, it will be important to validate and expand on the themes set out below with researchers and government users in other sectors.

Consultation Themes

Single Point of Service

Organizations such as Population Data BC (www.popdatabc.com) provide researchers with access to a range of data sets along with a variety of tools and services to support researcher data access and analysis (see Appendix B - Jurisdictional Summaries for more information on these services). However, researchers who are interested in working with data not held within such environments must work directly with each data provider to determine what data sets are available, apply for access and obtain data extracts (which are then linked by the researchers themselves or by another party on their behalf).

The potential for a single point of service (or “one-stop-shop”) to support access requests, approvals, linkage and analysis for data from multiple sectors was highlighted as a priority in each of the consultation initiatives. In the OIPC consultation, stakeholders from all groups agreed that a one-stop-shop to request and retrieve the data was a priority. Stakeholders in the BCHRS consultation envisioned a single point of entry to a virtual “hub” to support evidence-informed decision-making. The SPOR consultation also supported this vision of a “hub”. It also noted that a central data platform could act as a clearing-house for information about clinical trials (working with the BC Clinical Research Infrastructure Network, (BCCRIN)).

Security, Privacy and Governance

Across all of the consultation initiatives, there was recognition that effective security and privacy management are fundamental to the successful development and operation of facilities that store and manage citizen data. The SPOR, BCHRS and OIPC consultations highlighted the need for a secure data environment, recognizing that the abilities to secure data and protect citizen privacy are prerequisites for any such environment. During the Health Data Research Forum convened by the OIPC, it was acknowledged that strong foundations are in place to support the secure sharing of health information, but that significant opportunities for improvement still exist with respect to timely and appropriate access to this information for research purposes. Several recommendations arose from this meeting, including the development of proportionate and tailored privacy and security frameworks, clarification of roles and responsibilities, increasing standardization of processes and evaluating the option to harmonize processes across stakeholders where possible. This session also highlighted the need for a cross-sector governance framework to support clarity regarding accountabilities, consistency of decision-making processes and to enable the efficient and appropriate sharing of information.

Standardized and Transparent Processes

Provincial standardization or harmonization of processes for data access was an interest identified across all consultations. It was noted that there are currently numerous related and overlapping processes associated with gaining approved access to data (these include data access request processes, ethics approvals, grant applications and privacy impact assessments). As the timing of these processes is not aligned, and the requirements are often overlapping, this can result in significant duplication of effort for researchers and long delays in gaining access to data. It was noted that improved alignment could improve consistency in decision-making, reduce turnaround times and provide greater certainty, all of which could potentially support improved access to research funding. The role of governance was highlighted as being critical to achieve these objectives.

Improve Knowledge Translation

A focus on translating knowledge from research to practical application was highlighted across all three initiatives. Stakeholders in both the OIPC and BCHRS consultations expressed a general interest in a solution that builds competencies for knowledge translation. The SPOR consultation also identified this need, and noted the potential value of a “hub or bridging concept” to support improved interaction between researchers, policy makers and health care professionals to support alignment and more effective knowledge translation. Such a hub could also build capacity and capabilities related to implementation science (the study of methods to enable the integration of research results into healthcare policy and practice).

Build on Existing Platforms and Capabilities

Both the SPOR and OIPC consultations noted the opportunity to leverage existing infrastructure, capabilities and processes to enable more effective and efficient information sharing for health research purposes.

Enable Access to a Wider Range of Data Sets

The availability of comprehensive datasets was a common theme across all of the consultation initiatives. The ability to access and link a much broader range of health data, and across health and other sectors was identified as a high priority. University stakeholders in the OIPC consultation expressed the need for data access and linkages from multiple sectors, while Health Authority stakeholders were seeking decentralized access to real-time data across health authorities.

While other specific interests were identified throughout these consultation processes, the themes above were consistent and were important inputs for the development of the vision and value proposition. They also informed the development of the conceptual model, the components of which were selected to address many of the interests summarized above.

Relevant Trends








The Working Group conducted a scan of similar facilities to assess how other jurisdictions have approached the opportunities related to the use of linkable, public sector data by researchers and other users. The purpose of this section is to provide an overview of key trends identified through this scan, beginning with an overview of the facilities included in the review. This review was conducted using publicly available information and did not involve any interviews or validation with representatives of these facilities. This section is intended as a summary only, and readers are referred to the websites provided for each organization for additional detail (see Appendix B - Jurisdictional Summaries).




Organization Overview

The Working Group identified ten organizations/facilities to highlight environments that support linkage of government and/or health data for use by researchers and other user groups. Two private health systems in the United States were also included to provide a reference for advanced analytics in healthcare.

It is important to note that this scan was not intended to be exhaustive and that there are other organizations pursuing data linkage within the health sector and other sectors that have not been included here. The goal was to select sites from different regions and to describe different approaches while keeping the scan to a reasonable size. The following table provides a brief description of the organizations included in the scan.

Table 1 - Overview Of Organizations Included In Jurisdictional Scan

Jurisdiction	Organization	Summary
BC	 populationdata+	Population Data BC - Population Data BC is a multi-university, data and education resource facilitating interdisciplinary research on the determinants of human health, well-being and development
Canada	 Statistics Canada	Statistics Canada and McMaster - Research Data Centre (RDC) - Statistics Canada hosts RDCs at twenty-five universities across Canada that provide researchers with access, in a secure physical setting, to microdata from population and household surveys.
	 Manitoba Centre for Health Policy	Manitoba Centre for Health Policy (MCHP) - MCHP is a research unit in the University of Manitoba's Faculty of Medicine. MCHP acts as a steward for the information housed in the Population Health Research Data Repository on behalf of the Province for use by the local, national and international research community.
	 ICES	Institute for Clinical Evaluative Sciences (ICES) - ICES is an independent, non-profit organization whose core business is to conduct research that contributes to the effectiveness, quality, equity and efficiency of health care and health services in Ontario.
United Kingdom	 E-S-R-C	UK Administrative Data Research Network (ADRN) - The Economic & Social Research Council (ESRC) is establishing four national facilities to facilitate data linkage, analysis and access for broad public sector administrative data.
	 hscic Health & Social Care Information Centre	UK Health and Social Care Information Centre (HSCIC) - The HSCIC is a national provider of high-quality information, data and IT systems for health and social care in the United Kingdom.
	 CPRD	UK Clinical Practice Research Datalink (CPRD) - The NHS National Institute for Health Research (NIHR) has developed the CPRD to provide access to anonymized, linkable clinical data for clinicians and researchers.

Jurisdiction	Organization	Summary
Australia	 PHRN	AU Population Health Research Network (PHRN) - The Australian Government initiated PHRN as a data linkage network to enable health care planners and policy makers to respond to the changing needs of the Australian population.
United States	 Intermountain Healthcare	Intermountain Healthcare - Intermountain is a non-profit health system based in Salt Lake City, Utah, with 22 hospitals, over 800 physicians in the Intermountain Medical Group and a broad range of clinics and services.
	 Mount Sinai	Mount Sinai Hospital (NY) - One of the oldest and largest teaching hospitals in the United States. The hospital has a partnership with Continuum Health Partners in the creation of The Mount Sinai Health System, which encompasses the Icahn School of Medicine at Mount Sinai and seven hospital campuses in the New York metropolitan area.

Overview Of Key Trends And Observations

This section provides a summary of the trends and observations identified through a review of the above-noted organizations. For each category, the trend or observation is summarized and then examples are provided based on practices identified through the scan. The examples are intended to be illustrative, but not exhaustive, so examples are not provided for every organization in each case. Additional details regarding each organization and the online sources for all information below are available in Appendix B.



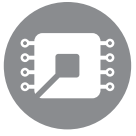
User Community

Trend: In most jurisdictions, the primary users are academic researchers and members of government, although permitting private sector users is an emerging trend.

Across the organizations investigated, the academic community and government (to support policy-focused research and to inform decision-making) were the primary users for these centres. In some cases, the user community also included private sector organizations; however, this is an emerging trend

Jurisdictional Examples

- Researchers represent the majority of the users across the facilities described above.
- Both the Health and Social Care Information Centre and the Clinical Practice Research Datalink engage with the pharmaceutical industry as a private sector user community. The Health and Social Care Information Centre has an indirect engagement model with anonymized data being provided to a non-profit organization i4Health, which then disseminates the data to pharmaceutical companies. i4Health has its own governance and services, including data analytics.
- Mount Sinai and Intermountain Health are the only private sector organizations included in this jurisdictional scan. In both cases, their primary user community consists of internal researchers, practitioners, and patients.



Data

Trend: Research centres have traditionally focused on administrative health data sets, but are working to expand to other sectors and types of data.

Some of the research centres in the jurisdictional scan primarily store and link administrative health data sets. In some cases, these are augmented by data from other sectors (e.g., social, economic or educational data) to support health research. Similarly, there is a trend to expand the variety and complexity of data to include additional sectors to support more advanced research.

Jurisdictional Examples

- While the Administrative Data Research Network supports health research, it is designed to enable access across a wide range of public sector administrative data sets. Cross-sector studies are currently one-time efforts that require additional legal approval, such as the UK study of the employment and benefit outcomes for ex-offenders. Research centres such as those within the Administrative Data Research Network aim to increase the frequency in which these cross-sector studies can take place. The value driven from cross-sector studies could include:
 - Addressing social mobility.
 - Researching causal pathways over a person's life (e.g., investigating links between social support and health).
 - Comparative analysis of access to and the provision of social care
 - Exploring the societal outcomes of mental health.
- Mount Sinai is one example of a research centre that is collecting data of greater variety and volumes in order to perform advanced analytics to support everything from personalized medicine to advanced strategies to enable operational effectiveness within their organization.



Data Access

Trend: Jurisdictions offer multiple channels for accessing data, with a growing trend toward enabling secure, remote access.

Many centres offer secure physical access at their facilities (which are often located at educational institutions). There is also a strong trend towards providing users with secure, remote access to their approved data sets. This approach provides flexibility to users while ensuring data sets are accessed and analysed in secure, controlled environments. Typically, data can only be removed once reviewed and approved through a formal process. Tools are provided within these virtual environments to enable analysis and documentation of results.

Jurisdictional Examples

- The Administrative Data Research Network is an example of leveraging existing research facilities. The four locations of the network are at universities across the UK, with each facility enhanced to create a secure research environment. In addition, the Administrative Data Research Network plans to establish secure rooms at higher education facilities across the UK for remote access for researchers.
- The Public Health Research Network is one example of a network that either enables researchers to link data themselves or offers remote access to a secure environment. The Secure Unified Research Environment (SURE) is a remote-access computing environment that allows researchers to remotely access and analyze the approved data extracts provided by data custodians for their research project. Researchers access their data through a virtual machine on their local computer screen. Within the SURE each researcher is allocated a virtual computer which runs entirely on hardware physically located at and controlled by the SURE. The SURE also contains extensive data storage and backup capabilities and a range of analysis software for researchers to use.
- Population Data BC provides access to approved data in a secure physical environment and via secure, remote connection into the “Secure Research Environment” (SRE). The SRE enables researchers to access and work with data remotely, but does not permit users to extract data. Research outputs can be extracted, but only after going through a formal review and approval process.
- The Institute for Clinical and Evaluative Sciences also enables remote access by uploading project-specific data extracts to a virtual environment that is accessible via a secure portal. All analysis of data is conducted within this virtual environment, and the outputs are released after a privacy and security review.
- The primary access method for Statistics Canada RDCs is at a physical facility (although some functionality is enabled through remote access as well). Once granted access, researchers have direct access to data sets through an on-site secure network environment with multiple security layers. Each workstation’s computer is connected to a central server to create a network. The network has no external connectivity (i.e. this network is isolated and is not connected to the internet or any other network).



Accreditation

Trend: Most jurisdictions require a minimum level of qualification and/or some form of accreditation for users.

In most cases, a qualification process is in place for users, and in some cases, formal accreditation is required (or is being proposed). The criteria differed between jurisdictions, but the common themes were that the researcher must demonstrate responsibility in privacy and ethics, as well as professional competency.

Jurisdictional Examples

- At the Manitoba Centre for Health Policy, new users are provided with an overview of the organization and the data access and use process. This includes their mission, available data and requirements for data use and publication of results. The process includes signing a pledge of confidentiality as well as an agreement that you understand and will follow the process for doing research at the centre.
- Statistics Canada requires that all users be able to demonstrate research experience and technical competence to work with the data requested. They also require that all users be named “deemed employees” of Statistics Canada, which results in them being subject to the same enforcement and compliance provisions as employees.
- At the Institute for Clinical and Evaluative Sciences, a researcher is accredited and given the title of an “ICES Scientist”, which is a probationary appointment. Accreditation is contingent on:
 - A University academic appointment at the level of Assistant Professor or higher;
 - A publication using ICES data;
 - Funds to carry out the proposed program of research; and
 - Absence of unmanageable conflicts of interest as determined by ICES.
- Accreditation at the Health and Social Care Information Centre requires the researcher to be able to demonstrate, to the satisfaction of the National Statistician, that he/she:
 - Has the appropriate knowledge and experience necessary for handling potentially disclosive personal information;
 - Has provided satisfactory evidence supporting their application that illustrates their professionalism and technical competence to carry out the research proposal; and
 - Demonstrates a commitment to protecting and maintaining the confidentiality of the data during the creation of outputs and publications that arise during the proposal.
- The Administrative Data Research Network is seeking to standardize accreditation across the UK for researchers. This accreditation process will require short course training including: best practice methods for working in a secure data environment; legal and ethical issues related to the use of individual records; methods of disclosure control; and an introduction to the variety of types of routinely collected administrative data. To retain accreditation status, researchers may be required to undertake an online update course on a regular basis.



Access Requests

Trend: Access requests and approval processes were similar across jurisdictions, although turnaround times varied.

All organizations included in this jurisdictional scan require a formal access request outlining the research to be completed, the rationale for the request and the data required. In most jurisdictions, the data centres provide support services to facilitate the drafting, review and approval of these requests. The time taken for fulfilling access requests varied considerably across the different research centres.

Jurisdictional Examples

- At the low end of the spectrum, access requests for answers to high level administrative/policy questions in some jurisdictions can be answered in as few as five business days. In contrast, more comprehensive access requests can take several months before approval.
- Both the Institute for Clinical and Evaluative Sciences and Public Health Research Network require researchers to justify that each element of information requested is absolutely necessary for their research. Once approved, researchers have access to their specific data sets within a virtual environment. In addition, the Institute for Clinical and Evaluative Sciences has a program that accepts research questions and responds with its own research.
- Population Data BC has a rigorous access request process involving submission of the request, data steward review, researcher contract agreement, and then provisioning of the approved data.
- Statistics Canada evaluates access requests for academic and scientific merit through a peer review committee, and determines what level of information is required to conduct the research. Once access is granted, researchers have access to a physical environment where they conduct their research.



Services

Trend: A wide range of services enable research centres to support users and data providers.

Services provided to users and data providers are a fundamental component of each of the organizations included in the scan. The types of services offered range widely from linking data sets, education regarding research methodologies and standards, to negotiation of contracts. Training is also provided and is offered through many channels (web-based e-learning, in-person workshops) and covers topics such as data privacy, ethics, available data sets, and data analysis.

Jurisdictional Examples

- Mount Sinai offers a secure Web-based learning module that provides investigators with the capability to manage training for their research projects to support compliance with protocols and data management.
 - The Institute for Clinical and Evaluative Sciences offers training to its internal employees in privacy policies and procedures relevant to their role.
 - Examples of expert research support include the client services team at the Public Health Research Network who can assist users with various aspects of projects involving linked data, study design and cohort selection, and completion of data applications. The Clinical Practice Research Datalink also follows the model of having an internal team of experts who provide services related to pharmacoepidemiology and pharmacoconomics.
 - Statistics Canada also offers support through analysts that are available to assist researchers on conceptual, methodological and procedural questions and on data concepts and data structures. Analysts can also direct questions to other Statistics Canada employees, if needed.
 - The Manitoba Centre for Health Policy has online research tutorials for database and research skills. In addition, the centre provides various courses to support health services and population-based research that is made possible using administrative data banks.
- An on-site Researcher will provide a user with start-up information on network/system access, space assignment (virtual and physical), system libraries and project access controls.
- Centres within the Administrative Data Research Network will offer short course training including: best practice methods for working in a secure data environment; legal and ethical issues related to the use of individual records; methods of disclosure control; and an introduction to the variety of types of routinely collected administrative data.
 - The Health and Social Care Information Centre has a range of services to help data users access, link and create custom data in bespoke formats. Data users can also call on the expertise and experience of data professionals to feed into their information projects such as Clinical Audit Management Service, Data Linkage and Extract Service, Secure Electronic File Transfer service and the NHS Safety Thermometer.
 - Population Data BC has a cost recovery model and charges researchers accessing its user-facing services. In addition to user-facing services, Population Data BC also offers services to data providers. These services include acting as a trusted third party for data linkage, managing distribution of data, and secure storage, archival and access of data. Additionally, Population Data BC provides training and offers a certificate program in Population Health Data Analysis for users.



Data Processing

Trend: Most jurisdictions separate identifying information from content prior to storage and linkage, and retain copies of original data sets centrally.

In most of the jurisdictions investigated, personally identifiable information is separated from “content” data physically and logically. Identifying information is replaced with random identifiers to support linkage. Most of the jurisdictions operate a “repository” model, meaning data sets are collected from Data Providers and stored centrally. Typically, there is minimal transformation or normalization of collected data sets, and they are used “as-is” to create approved extracts of linked data for researchers.

Jurisdictional Examples

- The majority of research centres link data sets internally.
- The Administrative Data Research Network uses an external party for all linking; record identifiers are matched by a trusted third party when data set linking is required, removing the need to transfer data with personally identifiable information to the centres. In contrast to this, the Clinical Practice Research Datalink only uses a trusted third party for linkage of externally held data sets such as ALSPAC and The Cancer Registry.
- The Institute for Clinical and Evaluative Sciences stores their data in an unlinked fashion, and links data sets as required by approved projects.
- The Public Health Research Network stores personal and content information separately and enables linkage of content information based on randomly-generated linkage keys



Compliance Monitoring & Enforcement

Trend: The approach to compliance monitoring and enforcement varied between jurisdictions, but most utilize a “right to audit” mechanism to assess compliance, with consequences for non-compliance defined contractually

Across the jurisdictions investigated, there was not a great deal of publicly-available information available regarding compliance monitoring and enforcement. A focus was observed on the approval mechanisms to grant parties access to data, and the right of the organization to audit any activities undertaken by users.

Jurisdictional Examples

- Population Data BC has security and compliance monitoring requirements as a standard part of their agreements and engages in monitoring of physical and logical access as well as regular audits.
- The Health and Social Care Information Centre requires all parties accessing their network to complete an Information Governance Statement of Compliance, which requires that no information is stored or processed offshore, that the party reports security events and incidents, and that the centre has a right to audit as required.
- The repercussions for breaching the terms of contracts can be quite severe. For example, consequences for the Public Health Research Network include:
 - Human Research Ethics Committee (HREC) approval suspended or withdrawn;
 - Data custodian refusing to provide data to the researcher in the future;
 - Data custodian refusing to provide data to researcher institution in the future;
 - Legal action over the breach of contract; and
 - Criminal prosecution which can result in fines or imprisonment.
- Population Data BC, the Clinical Practice Research Datalink the Public Health Research Network include contractual clauses that allow them to audit any researcher using their data in order to ensure compliance.



Big Data

Trend: Big data is an unexplored or emerging area in most jurisdictions

The use of big data is relatively new across the jurisdictions investigated. As more complex data is collected, the unique abilities of big data are being brought to bear on problems such as genomics. The ability to capture multiple formats and volumes of data has the potential to deliver value from a much wider range of data sources and to deliver new insights to all users.

Jurisdictional Examples

- At Intermountain, the Homer Warner Center for Informatics Research is involved in research into Clinical Decision Support, Data Modeling, Data Mining, Artificial Intelligence and applications in Clinical Genetics.
- Mount Sinai uses data analytics as an analytical, predictive and modelling tool to improve patient healthcare. For example, operational data from its EHR system is used to treat patients every day using a predictive model that analyzes clinical variables and generates the likelihood of 90-day readmissions among patients. Other uses of data analytics include:
 - Telemetry data is pulled from more than 1,000 beds and used for alerting
 - Development of a predictive model that analyzes variables associated with hospital acquired infections, which affect 1 out of 20 hospitalized patients
 - Development of risk models, powered by genomics, lab tests, billing records, and demographics, make up-to-date predictions about the individual patient a doctor is scheduled with.



Public Engagement

Trend: Many jurisdictions have developed strategies to engage the public in order to share the value of research on linkable, personal information and to solicit input on priorities.

Several jurisdictions have focused programs to engage the public for their input into the direction of research centres, and to provide information regarding the need for such research and the impacts it has on policy, decision making and clinical practice.

Jurisdictional Examples

- The Public Health Research Network has developed key strategies and programs to ensure that the ongoing development of the network is done with understanding and input from health consumer groups and the wider community. For example, each participant is required to have consumer and community participation in their development and researcher project planning as outlined in the network's Funding Agreement.
- The Clinical Practice Research Datalink runs a Patient and Public Involvement (PPI) program to gain input on what research should be conducted, how research can best be undertaken, and ways researchers should work with patients in research projects. Lay members are also a part of many committees that oversee the organization.
- The Administrative Data Research Network is planning on implementing a public engagement strategy to raise public awareness for the need of research based on linkable, anonymized administrative data, to engage the public through lay members on the Governing Board, and by holding public events.

Conceptual Model

In order to define options for the BC CDI in terms of technology, operating model and governance, a conceptual model was identified that functionally aligns with the vision for the project. This model is described in Figure 3 below, and contains seven key components, and two cross-cutting themes:

- Governance
- Data providers
- Data sources
- Data collection and aggregation
- Data storage and processing
- Reporting, analytics and visualization
- Customer functions (services)
- Data Consumers
- Core functions

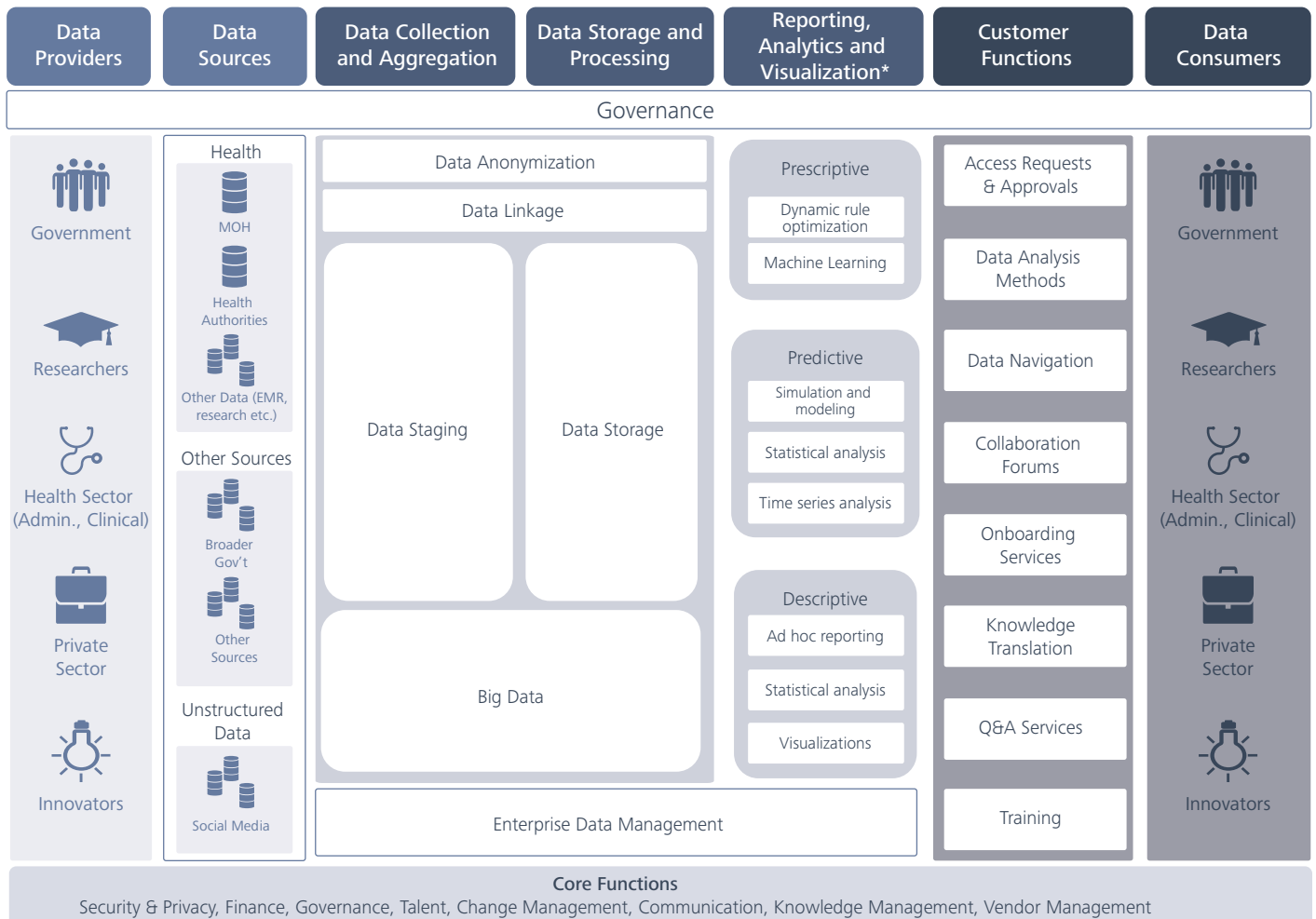
Each of these components and themes is described in greater detail below. This includes an overview of the component and its relevance to the vision of the BC CDI.

Overview Of Key Components Of Conceptual Model

Governance

A clearly defined governance model is critical to the success of the BC CDI as it will define the roles, responsibilities and decision-making processes necessary to guide the activities of the organization. To support its cross government vision and function, the governance model for the BC CDI should include representatives from sectors across government. The governance model could also include representatives from organizations fulfilling key roles within the conceptual model.

Figure 3 - Conceptual Model



Data Providers & Data Sources

Data Providers are the organizations that act as custodians or managers of data across a data domain (e.g., health, education, social services, etc.). Data providers are fundamental and necessary to enable the vision for the centre. As the custodians of the data within their respective sectors, Data Providers may manage numerous Data Sources (the systems used to store and maintain the data) and, they are familiar with the lineage, quality and characteristics of the data, as well as the security and privacy requirements associated with different data sets.

Data Collection and Aggregation

Data Collection is the process through which data is received from Data Sources and loaded into the BC CDI environment. There are numerous approaches to collecting data from source systems, including:

- Extract Transform and Load (ETL) - provides batch and real time data services that extract information from a data source, transform the data through a number of business rules, and load the data into a target data repository.
- Enterprise Application Integration - provides data services with pre-defined integration points both from a data structure and data access perspective.
- Data Virtualization - provides data services that integrate across varied data sources in real time, near real time, or batch. The virtualization layer acts as an intermediary to transform, and/ or cleanse data before data consumers access the data.
- Streaming - provides access to data sources which provide high-velocity data in real time.
- Big Data methods - these center on distributed processing frameworks leveraging the MapReduce model and dividing load across two or more servers. Big Data is often used as an umbrella term to represent high volume, velocity and/or variety of information assets that demand cost-effective, innovative forms of information storage and processing that enable enhanced insight, decision making, and process automation.

The type of data collection methods employed by the centre may vary depending on the data source and/or the functional requirements of the different user groups. For example, the application of ETL would be appropriate in situations where daily (or less frequent) data refresh is sufficient and the volume of data is low. On the other hand, the application of streaming or big data approaches would be more common for large volumes of data or highly variable data sets.

Data aggregation refers to an approach to standardization and integration that will enable data linkage across multiple data sources. It should be noted that the handling of data once acquired by the BC CDI could vary significantly. Data may be acquired and stored as received (once identifying information is removed), or it may be transformed to align to a common data standard to support improved integration (where permissible) in a data warehouse environment (or similar). Most of the jurisdictions surveyed (see "Relevant Trends") applied minimal transformation to incoming data sets. Rather, data sets were stored as received (once identifiable information was removed) and then linked on an as-needed basis for approved purposes.

Data Storage And Processing

Data Storage refers to the mechanism by which data is stored and made available for analysis purposes. This could include approaches utilizing traditional data warehouse architecture and/or more advanced approaches involving distributed processing and storage. The mechanisms and infrastructure chosen for the centre will depend upon the data available and the requirements of users. It is likely that the centre will require both traditional and big-data capabilities to support the fundamental analytics needs of users as well as emerging analytics and data processing requirements in fields such as genomics research.

It is important to note that de-identification will be a critical process within the BC CDI. This process involves removing identifiable information from incoming data sets, while supporting linkage between these data sets in a consistent manner that addresses the security and privacy requirements of data providers and the BC CDI. There are numerous approaches to anonymization that focus on balancing the risk of re-identification against the utility required by users, and such services can either be delivered within the centre or by a trusted third party.

Reporting, Analytics and Visualization

The Reporting, Analytics, and Visualization component represents a suite of tools that enable varying levels of analysis of data within the centre. Analytics approaches typically utilize standard descriptive analysis (critically important for identifying trends in data and testing research hypotheses). More advanced predictive modeling uses available data to create models to predict outcomes for a specific scenario or range of conditions, while prescriptive capabilities build on predictive analytics by providing recommendations regarding actions to take given certain predicted outcomes.

Customer Functions

The customer functions represent the services provided to stakeholders engaged with the centre. The primary consumers of services for the centre will be the users of the data; however, it is important to note that there will likely be a critical set of services required to support data providers in order to encourage contributions of data to the environment, ensure that data is available in the format required and to ensure that all security and privacy requirements are understood, addressed and reported against on an ongoing basis.

Below is a brief overview of a range of relevant services that could be provided by the centre:

Access Requests & Approvals - This includes support in the definition phase, processing of all access requests, reporting on progress and finally reporting on the outcome of the request.

Data Analysis Methods - This includes subject matter experts who could provide guidance regarding approaches for study design and statistical analysis. The goal would be to promote consistency in definitions and methods among users to enable comparability of results.

Data Navigation - This includes providing services to data consumers to support awareness of available data sets and options for addressing specific questions or challenges based on the data available.

Collaboration Forums - This includes services to engage with Data Consumers to promote cross-organizational exchange and analysis of data.

Onboarding services - This includes resources to work with Data Providers to review potential data sets for acquisition by the BC CDI, define appropriate terms and conditions for sharing, protecting and monitoring their data and to define services required on behalf of the Data Provider for users.

Knowledge Translation - This includes service to make data findings and insights applicable in practice for Data Consumers and the general public.

Training - This includes services to provide training to Data Consumers.

Q&A Services - This includes support for questions from Data Consumers related to data and its use. This would include the ability to answer specific questions regarding individual data sets and variables and support in the interpretation of results.

Data Consumers

Data Consumers are those organizations or individuals who will utilize the data, infrastructure and services provided by the BC CDI. It is anticipated that the initial focus of the centre will be on Government, Health Sector and Researcher users. However, where possible and appropriate, additional user groups will be evaluated to ensure the greatest value is being derived from this critical information resource.

Core Functions

There are a number of core functions or services that will be required internally to support the ongoing operations of the BC CDI. These include standard corporate services such as finance, security, privacy, human resources, legal and communications as well as the capacity to oversee vendors and to initiate and deliver major projects related to technology and business processes. Additionally, it is proposed that the BC CDI have in-house compliance monitoring and enforcement capabilities to ensure constant oversight of operations, support continual improvement through ongoing audits and reviews and, as appropriate, to identify violations of policy or contract terms and follow up with appropriate enforcement measures.

Costs and Funding Options

Comparison Of High-Level Costs In Other Jurisdictions

To realize the vision, goals, and objectives of the BC CDI, it is important to understand the financial costs related to both developing a centre and maintaining it on an ongoing basis. The Working Group took a multi-step high level approach to arrive at a range of costs. The approach started by examining the costs of comparable organizations in other jurisdictions.

Table 2 below highlights the approximate budgets for similar facilities in other jurisdictions, where such data was publicly available. These figures are provided for reference, and an effort has been made to apportion budgets between capital and operating costs (where possible). It should be noted that there is significant variability in the scope, scale and capabilities of these facilities, and this results in a large range in both capital and operating budgets. Where possible, additional comments have been included to explain the figures provided and some of the key differences between these facilities.

Table 2 - Approximate investments in infrastructure and operating expenses for select jurisdictions (all figures converted to CAD\$ for comparison purposes)

Organization	Infrastructure Investment	Approximate annual operating costs	Comments
Statistics Canada Research Data Centres	Not available	~\$250K/year (per RDC) ²	Operating budget includes funding for space, staff, and updates and replacements to hardware and software
Manitoba Centre for Health Policy	\$3.5M (1990) ³	\$5.7M ⁴	\$5.7M annual operational funding (2012/2013) includes \$2.3M in provincial funding and \$3.4M in awards, grants, and contracts
Institute for Clinical Evaluative Sciences (ICES)	Not available	\$21.9M ⁵	<ul style="list-style-type: none"> On top of operating the infrastructure to support data management and linkage, ICES employs a team of scientists to support its mandate. Operating costs reflect expenditures for 2012/2013 A significant proportion of operating expenses is covered by research and operating grants
UK Administrative Data Research Network (ADRN)	\$63M for development and maintenance over 5 years		<ul style="list-style-type: none"> Funding is provided for four data centres and the Administrative Data Service. These centres will build upon existing technology and infrastructure where possible and the funding is in addition to other sources of funding currently in place.
Health and Social Care Information Centre (HSCIC)		\$85.8MM ⁶	<ul style="list-style-type: none"> 2011/2012 expenditures The HSCIC is a large organization (538 people as of 2012) and this figure includes staff and non-staff costs
AU Population Health Research Network (PHRN) ⁷	\$20.3MM (2006)	\$15.6MM	<ul style="list-style-type: none"> In 2006, the National Collaborative Research Infrastructure Strategy allocated \$20.3MM for population health data linkage. Annual operating expenditures for the network were for 2011/2012. Investment in the national network to date totalled \$71.4MM

Reference:

2. <http://www.statcan.gc.ca/rdc-cdr/faq-eng.htm>

3. <http://www.longwoods.com/content/22117>

4. MCHP Annual Report (2012/2013): http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departamental_units/mchp/media/MCHP_2012_Annual_Report.pdf

5. ICES Annual Report (2012/2013)

6. HSCIC Annual Report and Accounts (2011/2012): <http://www.official-documents.gov.uk/document/hc1213/hc03/0332/0332.pdf>

7. PHRN Annual Report (2011/2012): http://www.phrn.org.au/media/79221/annual%20report%202011_2012v1.0.pdf

Preliminary Cost Estimates And Assumptions

Following the jurisdictional scan, a preliminary cost estimate was created based on the conceptual model framework illustrated above. The scenarios that were considered include the costs to create a net new facility to enable BC CDI, and the costs to support expansion when adding data providers or entire sectors.

To align with the conceptual model, the cost estimate is broken up into the following major buckets: Governance, Data Collection and Storage, Enterprise Data Management, Reporting, Analytics, and Visualization, Customer Functions, Physical Infrastructure, and Core Functions. The following high-level assumptions were made in these areas:

Governance - Executive Governance will be added as needed and will leverage existing sector / cross-sector vehicles. Governance roles and responsibilities will be provided by key leadership roles at BC CDI that encompass Executive, Operational, and Technology/Data Governance. The leadership roles at BC CDI will not only have these governance roles, but additional roles and responsibilities within the centre. In addition to Executive Governance, governance will need to be provided at an operational level. (e.g., shepherding data requests through the approval process)

Data Providers and Sources - The cost estimates only include costs for establishing and operating the centre and do not include the participation costs for Data Providers. These costs will vary significantly across providers depending on the scale, complexity and current state of their data holdings.

Data Collection and Storage - This model assumes a data environment where incoming data sets are not normalized or integrated into a data warehouse. Data sets are temporarily stored separately as flat files and when data requests are submitted (and approved), there will be on-demand linkage across these flat files. The key data acquisition services of anonymization and linkage are assumed to take place centrally and are driven by the number of data domains (e.g., Patient, Provider) requiring anonymization or linkage. Finally, data storage costs were allocated for both data acquisition (on demand) and for researcher usage.

Enterprise Data Management - The model assumes centralized roles for the following functions: Data Privacy and Security, Data Architecture, Data Retention and Archiving, Metadata Management, Master Data Management, and Data Quality Management. The majority of the resources will be in the areas of Data Privacy and Security, Master Data Management, and Data Architecture with some allocation for the other areas.

Reporting, Analytics, and Visualization - The model only assumes the costs for acquiring Enterprise licenses of common analytical tools for researcher consumption. The costs for developing any artifacts or outputs from these tools are assumed to be incurred by the Data Consumer leveraging these tools.

Customer Functions - The model assumes centralized roles for all the functions with particular support for Access Requests and Approvals, Q&A Services, Knowledge Translation, and Training. A significant proportion of the costs are allocated to Access Requests and Approvals due to the fact that data sets are not normalized thus resulting in manual effort to create unique data sets for each data request.

Physical Infrastructure - The model assumes costs for IT infrastructure and for office space to support the delivery of the above components. The components of IT infrastructure include physical device costs and operating costs for Security, Management and Utilities (i.e., Power and Cooling).

Core Functions - The model assumes costs for key support roles including: Finance, HR, Legal, Privacy, Operations, and Communication functions.

General assumptions also include the following:

- Costs incurred at the center were included in the estimate with participant costs of data providers or sectors not included.
- Average staff costs (including salaries and benefits) were used.
- Hourly rate with assumed IT implementation costs (i.e., assume effort for requirements/analysis and forecast costs on future phases based on a percentage of requirements/analysis).

Based on the assumptions above, the costs to stand up a net new facility would range from \$7.0MM - \$10.5MM for capital expense and \$4.0 MM - \$6.0 MM for operating expense. The costs above have a range based on some considerations which include being able to leverage or accelerate the setup of capabilities based on existing infrastructure or capabilities of a proponent who has carried out similar initiatives.

It is difficult to estimate the additional capital and operating costs associated with onboarding new data sets or sectors, due to the significant variability in scale, data quality, data format and data types. Additional costs were estimated, but varied widely due to this range in complexity. For example, data sets that are not able to be linked through a consistent key would require additional effort to link consistently and completely. In addition, the format in which data sets are received (e.g., flat file or unstructured) and the degree of standardization across different data providers would drive up the incremental costs. The range of costs for incorporating new data sets include capital costs ranging from 10% - 90% of the costs to stand up a net new facility (with the lower range associated with on-boarding a relatively small data set and the higher range associated with on-boarding an entirely new sector). The range of costs to operate a facility with an additional data provider or sector will range from 10% - 100% of a net new facility.

Funding Options

A key challenge in the development and operation of a centre like the BC CDI is to establish sustainable, long-term funding to cover capital and ongoing operational costs. As highlighted above, the scale of these costs is highly dependent upon the scope and services of the centre, with costs being proportional to the number of data sets and users involved. Five potential sources of funding have been identified for the BC CDI below.

Government Funding

Federal and Provincial agencies provide funding for initiatives similar to the BC CDI across Canada. Many of the organizations reviewed in the jurisdictional scan received a portion of their funding (either one-time capital funding or ongoing operational funding) through such government agencies, and similar opportunities exist in BC. For example, CIHR's SPOR funding could potentially be applied to fund components of the BC CDI that are specifically focused on patient-oriented health research. Similarly, organizations such as the Canadian Foundation for Innovation (CFI) and BC's Knowledge Development Fund (KDF) provide funding to support research infrastructure with the goal of ensuring that maximum benefits are realized from these investments.

As the BC CDI launches with a focus on health data, it could leverage investments from health funding agencies (such as CIHR's SPOR strategy) and from CFI and KDF. As additional sectors are incorporated into the BC CDI, the potential exists to attract similar funding from these sectors, thus aligning funding from multiple sources for a similar purpose, and reducing the potential for duplicated infrastructure and investment across sectors.

Research Grant Funding

In some of the jurisdictions assessed, ongoing operational costs were covered in part by project-specific research grants. Such funds are typically provided by research funding agencies and are awarded on a competitive basis (typically annually or on a multi-year basis).

Private Sector Partnerships

Another potential opportunity to offset some of the capital and/or operational costs for the BC CDI is to seek out private sector partners who are willing to provide services and/or infrastructure in support of the centre on a reduced-cost, cost-sharing or pro-bono basis. Such an arrangement may also enable these partners to develop innovative products and/or services to support users of the centre.

Data Provider Fees

Centres such as the BC CDI can provide services to researchers on behalf of the data custodians who are making data available for research purposes. In exchange for providing these services, the Data Providers, realizing administrative savings by delegating these services to a third party, could provide financial support to the BC CDI.

Data Consumer Fees

Finally, many of the organizations reviewed in the jurisdictional scan used a fee-for-service model to recover a portion of the costs for creating and operating their infrastructure and for the services and tools they provide.

Recommendations

Recommendations

This section outlines eight recommendations that, if approved, will enable short-term benefits for the health research community while building a foundation for the broader BC CDI vision.

1 Approve the Vision and Value Proposition for the BC CDI and initiate a multi-year, phased implementation approach to realize this vision

Based on an analysis of trends in other jurisdictions, the interests of stakeholders within BC and the potential benefits that the BC CDI could deliver to British Columbians, it is recommended that the long-term Vision and Value Proposition be approved as a starting point for a multi-year, phased approach to creating a BC Centre for Data Innovation.

The recommendations that follow propose Phase 1 activities or next steps to support this phased implementation approach.

2 Establish an Interim Governance Committee to oversee the ongoing development of the BC CDI

It is recommended that an Interim Governance Committee be established to oversee and guide the implementation of the recommendations set out in this report. This committee should include representation from the Ministry of Health and the Ministry of Technology, Innovation and Citizens' Services, as the current sponsors of this initiative, but should also include representation from other sectors to ensure a cross-government approach and perspective.

As the BC CDI expands to include data sets from other sectors and a broader range of data providers and data consumers, this interim committee could be tasked with making recommendations for a permanent governance structure to oversee the operations and continued growth of the BC CDI.

3 Leverage existing capabilities and technology to build the foundation for the BC CDI, starting with the Ministry of Health

A key principle guiding the Working Group was to assess where it was possible and desirable to leverage existing infrastructure, capabilities and processes to accelerate the development of the BC CDI. It is recommended that the existing relationship between the Ministry of Health and Population Data BC be leveraged as a Phase 1 activity, and that the capabilities and processes currently in place be enhanced to improve researcher access to data held within the Ministry of Health's primary data warehouse (HealthIdeas). HealthIdeas contains a wide range of health data (including acute care, Medical Services Plan, home and community care and vital statistics data) which is invaluable to health-related research.

Implementation of this recommendation will not only provide more timely access to a range of current health data sets, but will also lay the foundation on which to build a cross-government data innovation centre and service. In particular, implementation of this recommendation will enable and test several of the fundamental components illustrated in the BC CDI conceptual model (see section entitled "Conceptual Model"), such as data storage and processing, reporting, analysis and visualization and a range of customer services.

More detailed information on this recommended Phase 1 activity, including specific next steps, can be found in Appendix C.

4 Broaden the scope of consultation and inquiry to include additional sectors beyond health and additional stakeholders/ data consumers beyond academic researchers

In assessing the potential for a provincial data innovation centre, the Working Group was asked to start with health sector data and interests and align with the work that was already underway within the health sector to improve access to data for academic researchers. While this work is by no means complete, implementation of Recommendation 3, and the supporting Health Vision outlined in Appendix C, will set the foundation for improved access to health data by academic researchers and government entities.

However, additional consultation, validation and inquiry will be necessary to ensure that the BC CDI is designed with broader public sector interests in mind and to serve users beyond the academic research community. It is therefore recommended that the Interim Governing Committee, as a first step, expand the current dialogue to include other government sectors and other potential user groups with a view to identifying additional partners and stakeholders with an interest in the BC CDI.

While the primary purpose of this consultation should be to work with other sectors to identify high-value data sets that could be made available through the BC CDI, the process could also be used to validate and further refine the vision and value proposition for the BC CDI, gather additional requirements and identify additional infrastructure, capabilities and processes that could be leveraged by the BC CDI.

5 Build robust security and privacy protection into each phase of the BC CDI implementation

Robust security and privacy protection are critical to the successful realization of the BC CDI vision. It is therefore recommended that a thorough assessment of security and privacy risks and opportunities be conducted for each implementation phase and whenever a new organization's data sets are made available through the BC CDI. Further, consultation with the Office of the Information and Privacy Commissioner during each phase is recommended to ensure that privacy and security concerns and risks are identified and addressed and that acceptable controls and safeguards are put in place to maintain public trust in the BC CDI.

6 Streamline the Access Request and Approval Process

Streamlining the access request and approval process through which researchers and other users currently obtain access to health data was a recurring theme raised by the majority of stakeholders (see "Stakeholder Interests"). For the BC CDI to achieve the identified benefits, it must not only be trusted to enable access to high-value data sets with a high degree of privacy and security, but it must do so in a timely and reliable manner. The current process which involves, in some cases, multiple levels of approval and unpredictable delays must be rectified before it can be leveraged and expanded to include broader public sector data and providers.

To improve the current process and to position the BC CDI as a world class centre for data access and innovation, it is recommended that the Interim Governance Committee assemble and oversee a multi-stakeholder team comprised of government representatives and users of the service to develop practical recommendations for streamlining and harmonizing the current access request and approval process. These recommendations should be developed with a view to supporting broader government and user interests. As a starting point, it is recommended that the team consider the findings and recommendations contained in the recent reports arising from the meetings convened by the Office of the Information and Privacy Commissioner (Report of the Roundtable Discussion on Access to Data for Health Research, and Report of the Health Data Research Forum^{8,9})

Reference:

8. Office of the Information and Privacy Commissioner. 2012. Report of the Roundtable Discussion on Access to Data for Health Research. 17 pp.

9. Office of the Information and Privacy Commissioner. 2013. Report of the Health Data Research Forum. 20 pp.

7 Explore the short-term opportunity to leverage available government funding for Phase 1 activities while developing a long-term, sustainable funding and revenue model

As set out in the section on funding options, a number of similar initiatives in other Canadian jurisdictions have been supported in part through government funding and grants. Similar opportunities exist for BC, such as funding which is currently available through CIHR's SPOR strategy, and it is recommended that the Ministry of Health explore whether CIHR funding could be used to support the Phase 1 activities proposed in Recommendation 3.

Ultimately, it is recommended that the Interim Governance Committee develop a long-term funding model that is both flexible (to include multiple sources of funding and revenue) and sustainable. Options for consideration, which are described in more detail in the section entitled "Costs and Funding Options", should include government funding, private sector partnerships, shared service fees and user fees.

8 Evaluate opportunities to leverage emerging technologies and advanced analytics capabilities to maximize the value of public sector data

The BC CDI will offer significant potential to evaluate and leverage emerging technologies and analytics capabilities in order to make the best use of the BC Government's vast data resources. For example, big data technologies could be used to support emerging research and medical care in the field of personalized medicine. These technologies enable analysis across a wide range of large, variable and complex data sets, and could result in new insights to support policy development, decision-making and innovation across a number of sectors.

However, such opportunities also raise concerns with respect to security and privacy, and must be carefully considered and prudently implemented. It is recommended that the opportunities and benefits posed by emerging technologies and advanced analytics capabilities be explored and carefully evaluated for privacy and security risk before being implemented in a broad way. A "proof of concept" approach may be ideal for exploring both the potential opportunities and unintended risks of these emerging technologies.

Appendix A - Overview of Consultation Initiatives

This section provides an overview of the consultation initiatives conducted previously within the health sector.

BC Health Research Strategy

The BC Health Research Strategy (BCHRS) is an initiative facilitated by the Michael Smith Foundation for Health Research to “shape a more comprehensive, coordinated, and systems-oriented approach to health research” in BC¹⁰. The process was initiated in the fall of 2012, with consultation taking place from March - July, 2013. The consultation process engaged health research and healthcare leaders from across the province. Initially, key informant interviews resulted in the identification of five “directions” for the strategy that then served as a framework for consultation. This was followed by a series of focus sessions (one session for each “direction”) designed to refine each direction, validate feasibility and gather additional advice regarding implementation considerations.

To enable input from the broad health research community in BC, an online survey was distributed to approximately 3,000 individuals to gather feedback regarding the draft vision and goals, potential provincial-level actions and key research program gaps that could be addressed by a provincial health research strategy. Five regional workshops were held to engage representatives from each geographic region to gather input and validate that the health research strategy would address each region’s particular requirements. Finally, a validation workshop was held to refine and prioritize the actions to be included in the final strategy for supporting high-quality health research¹¹.

Canadian Institutes for Health Research - Strategy for Patient Oriented Research Business Plan

The Canadian Institutes of Health Research (CIHR) has created a national “Strategy for Patient Oriented Research” (SPOR). This strategy aims to improve health outcomes and enhance patients’ health care experience through integration of evidence at all levels in the healthcare system. One of the key components of SPOR is the development of Support for People and Patient-Oriented Research and Trials (SUPPORT) Units. These units are proposed to function as research service centres that provide specialized expertise and infrastructure to support patient-focused research.

There are several such units across Canada, and a business plan is currently being created for a proposed unit in BC. One of the proposed “core functions” of the SUPPORT unit in BC is “Data Platforms and Services”. BC is launching a refreshed strategic vision for the provincial health system (“Setting Priorities for BC’s Health System”). This strategic vision will provide a focus to the SUPPORT unit.

To support the development of the business plan, consultations were held with BC health system leaders, health care providers, health researchers and patient representatives. Health system leaders were engaged in March, 2013 to identify critical functions that would increase the quality and productivity of research through the SUPPORT unit in BC. Consultations were then conducted with representatives of the research community to refine these priorities and inform the development of an Expression of Interest to identify individuals to support the development of the business plan. Ongoing engagement with these stakeholder groups continued into early 2014 as the business plan was developed and refined¹².

Office of the Information and Privacy Commissioner

The Office of the Information and Privacy Commissioner convened two meetings on the topic of access to data for the purposes of health research. The first session (Roundtable Discussion on Access to Data for Health Research) was held on June 25, 2012, and was convened by the Commissioner to: 1) “determine if there were...problems with the interpretation or administration of FIPPA” that may be impeding access to health data for medical research purposes, and 2) “to discuss the opportunities, barriers and possible solutions to improving access to data for health research”¹³.

The second session (the Health Data Research Forum) was held on December 9, 2013 and was designed to build upon the efforts and outputs of the first meeting. The forum was attended by representatives from the Office of the Information and Privacy Commissioner, the Ministry of Health, the academic research community, health authorities and individuals from various agencies, councils and other professional privacy organizations. Specifically, the purpose of the meeting was to “discuss and seek solutions on timely access to, and appropriate use of, health data under the stewardship of the MOH and Health authorities for research, evaluation, planning and quality improvement while maintaining academic independence and enhancing privacy protections and security safeguards”¹⁴.

Reference:

10. <http://bchealthresearchstrategy.ca/about/>

11. British Columbia Health Research Strategy - Validation Workshop Summary of Discussions. (http://healthresearchstrategy.files.wordpress.com/2013/11/bchrs-validation-workshop-report_final.pdf)

12. Michael Smith Foundation for Health Research. 2013. SPOR Support Unit Project - Consultation Sessions: Summary Report. http://www.msfr.org/sites/default/files/SPOR_consultation_summary_report.pdf

13. Office of the Information and Privacy Commissioner. 2012. Report of the Roundtable Discussion on Access to Data for Health Research. 17 pp.

14. Office of the Information and Privacy Commissioner. 2013. Report of the Health Data Research Forum. 20 pp.

Appendix B - Jurisdictional Summaries

Objectives

In order to inform the activities of the Working Group, a jurisdictional scan was undertaken to assess how other jurisdictions have approached the creation of similar environments. The goal was to identify leading practice, emerging trends and lessons learned based on the experiences in these jurisdictions.

This section provides a summary of the information collected during this scan and is provided for reference purposes. It should be noted that the information below is based entirely on publicly available information and has not been validated with representatives from these jurisdictions. The summary for each jurisdiction is not intended to be exhaustive, and readers are directed to the references cited for each jurisdiction for additional detail and the most up-to-date information.

Approach

The approach for the Jurisdictional Scan included the following characteristics:

- A selection of facilities that provide government data for research were identified in Canada and internationally (see Table 3 below).
- Candidates for the scan were selected to highlight different approaches to health sector and broader public sector data aggregation, linkage, access and analysis.
- Each centre was analyzed against a series of questions outlined in a framework that was created to guide consistency in the analysis (see section below).
- Information was collected through secondary research (primarily internet-based).
- One interview was conducted with Statistics Canada, to provide additional context to its operations and process.











Framework

A framework was developed to support consistent assessment of each of the organizations included in the scan. The Framework consists of six components, each of which included several questions to guide the research.

Each of the components of the model is defined below:

- Data Sources - includes an overview of sectors or organizations from which the facility obtains data.
- User Community - describes the stakeholder groups that utilize the centres and/or receive a service from the centre.
- Service - includes the services provided to all stakeholders that interact with the facility (including Data Providers and Data Consumers).
- Governance - provides an overview of the governance model, including people and organizations involved and their roles and responsibilities where available.
- Data Access and Approvals - describes the process by which users request and gain access to data, and the approval process for such requests.
- Security and Privacy - includes an overview to the approach taken to address security and privacy requirements, including alignment to regulation and policy

Table 3 - Overview Of Organizations Included In Jurisdictional Scan

Jurisdiction	Organization	Summary
BC	 populationdata	Population Data BC - Population Data BC is a multi-university, data and education resource facilitating interdisciplinary research on the determinants of human health, well-being and development
Canada	 Statistics Canada	Statistics Canada and McMaster - Research Data Centre (RDC) - Statistics Canada hosts RDCs at twenty-five universities across Canada that provide researchers with access, in a secure physical setting, to microdata from population and household surveys.
	 Manitoba Centre for Health Policy	Manitoba Centre for Health Policy (MCHP) - MCHP is a research unit in the University of Manitoba's Faculty of Medicine. MCHP acts as a steward for the information housed in the Population Health Research Data Repository on behalf of the Province for use by the local, national and international research community.
	 ICES	Institute for Clinical Evaluative Sciences (ICES) - ICES is an independent, non-profit organization whose core business is to conduct research that contributes to the effectiveness, quality, equity and efficiency of health care and health services in Ontario.
United Kingdom	 E-S-R-C ECONOMIC & SOCIAL RESEARCH COUNCIL	UK Administrative Data Research Network (ADRN) - The Economic & Social Research Council (ESRC) is establishing four national facilities to facilitate data linkage, analysis and access for broad public sector administrative data.
	 hscic Health & Social Care Information Centre	UK Health and Social Care Information Centre (HSCIC) - The HSCIC is a national provider of high-quality information, data and IT systems for health and social care in the United Kingdom.
	 CPRD	UK Clinical Practice Research Datalink (CPRD) - The NHS National Institute for Health Research (NIHR) has developed the CPRD to provide access to anonymized, linkable clinical data for clinicians and researchers.
Australia	 PHRN	AU Population Health Research Network (PHRN) - The Australian Government initiated PHRN as a data linkage network to enable health care planners and policy makers to respond to the changing needs of the Australian population.
United States	 Intermountain Healthcare	Intermountain Healthcare - Intermountain is a non-profit health system based in Salt Lake City, Utah, with 22 hospitals, over 800 physicians in the Intermountain Medical Group and a broad range of clinics and services.
	 Mount Sinai	Mount Sinai Hospital (NY) - One of the oldest and largest teaching hospitals in the United States. The hospital has a partnership with Continuum Health Partners in the creation of The Mount Sinai Health System, which encompasses the Icahn School of Medicine at Mount Sinai and seven hospital campuses in the New York metropolitan area.

Population Data BC

- Population Data BC is a university-based cross-sector research facility, offering comprehensive educational services, and focusing on the determinants of human health, well-being and development.
- Population Data BC supports:
 - **Data Access:** Population Data BC provides access to individual-level, de-identified longitudinal data on British Columbia’s residents. Linkage of data across sectors can help researchers understand the causalities in complex factors that span a person’s life, and can help inform policy-making and translate into healthier communities.
 - **Education and Training:** Educational services include certification for population health data analysis, in-person workshops, and free online courses.

Data Sources

- Datasets available at Population Data BC include:
 - Health care and health services data
 - Population and demographic data
 - Occupational data
 - Early childhood data
 - Spatial data
- Some additional health care and health services datasets are externally held and are linkable upon request

Users

- Researchers

Services

- Cross-sector dataset linkage services
- Expert guidance, support and advice regarding datasets
- Secure remote environment for data storage and analysis
- Educational services include certification for population health data analysis, in-person workshops, and free online courses
- Services for data providers include managing the data access request process, secure storage, handling, and archiving of data, acting as a trusted third party for dataset linkage, and providing a secure environment for researchers

Governance

- Population Data BC’s multi-tier governance framework includes the following:
 - The Data Stewards Working Group consists of Data Stewards from each data provider who ensure that the data access process meets their privacy and security expectations.

- The Advisory Board guides Population Data BC on core issues, such as strategy, policy, funding, and security. The Advisory Board is comprised of leaders in the academic research field.
- The Governance Oversight Committee guides and supports the operations of Population Data BC. The Governance Oversight Committee oversees the Advisory Board and is composed of VPs of Research at BC’s research universities, members of Population Data BC’s Advisory Board, and researchers.
- Population Data BC’s Executive Director and unit Leads are responsible for the operational management of Population Data BC.

Access and Approval Process

- Access is available to individual level de-identified linkable data on British Columbians. The access request process follows 7 stages:
 - Submitting a data access request
 - Data Steward review and approval
 - Privacy contract and researcher account set up
 - Data preparation and delivery
 - Data analysis
 - Publishing of approved research materials
 - Project closure

Security and Privacy

- Advanced physical, technical and procedural measures are in place to ensure individual privacy and data protection (the reader is referred to Population Data BC’s website for a detailed description of these controls under “Privacy” and “Data Linkage”)
- Population Data BC aims to meet and exceed all relevant legal, ethical and legislative guidelines for data protection and privacy

Reference:

<https://www.popdata.bc.ca/>

Statistics Canada RDC Network (est. 2000)

- Research Data Centres (RDCs) provide researchers with access to data from population and household surveys in a university environment
- There are 26 centres across Canada, 25 of which are based at a university and operated by Statistics Canada analysts. There is one centre in Ottawa to service the needs of the federal government
- The centres are operated under the Statistics Act and are accessible only to researchers with 'deemed employees' status, and who have been sworn in under the Statistics Act.

Data Sources

- Statistics Canada provides survey data across community, health, education, aboriginal, income, social network, diversity, economical, labor, family and ethnic diversity.
- Provides additional data sources with statistical publications, The Canadian International Merchandise Trade (CIMT) online database, Canadian and U.S. trade data at the 6-digit commodity level (HS) and CANSIM

Users

- Researchers (government and academic)
- Users outside of the provincial research data centre can access data with the establishment of the Federal Research Data Centre (FRDC), in Ottawa, in response to a growing demand for access to detailed micro data for analytical purposes in federal departments

Services

- Analysts are available to assist researchers on conceptual, methodological and procedural questions and on data concepts and data structures.
- Throughout the year, conferences are held by the research network that provide information sessions on data sets, statistical programs or procedures and methodological issues.

Governance

- Social Sciences and Humanities Research Council (SSHRC), Canadian Institutes of Health Research (CIHR), Canada Foundation for Innovation (CFI), Statistics Canada and a consortium of universities across the country are responsible for RDCs

- In most cases, the Academic Director is a faculty member at the hosting university and the primary liaison with the RDC Program and the hosting university. The Academic Directors' responsibilities include: promoting the RDC's research activities on campus and advocating for the local RDC within the RDC Program initiative.

Access and Approval Process

- Each researcher must sign the Microdata Research Contract (MRC)
- Each researcher is given a unique username and password by the Analyst to access the approved data set(s) and folder(s) for each contract
- Statistics Canada offers virtual access via the Real Time Remote Access (RTRA) system; an on-line remote access facility allowing users to run SAS programs, in real-time, against data sets located in a central and secure location.

Security and Privacy

- An RDC has a secure network environment with multiple security layers. Each computer is connected to a central server to create a network (which is isolated and is not connected to the internet or any other network).
- All analysis must be conducted at the RDC, with data stored on the RDC server. Drives/connections on the computers are disabled to protect against accidental breaches and harmful computer viruses. RDC Analysts review all data to be loaded to, or extracted from, the server.
- All research outputs are vetted before they leave RDCs (or any Statistics Canada offices). Microdata never leave the secure area of the RDC.

Reference:

Statistics Canada RDC Network: <http://www.Statistics Canada.gc.ca/rdc-cdr/index-eng.htm>
<http://www.Statistics Canada.gc.ca/rdc-cdr/pdf/report-rapport-eng.pdf>
 Research Data Centre Guide: <http://socserv.socStatistics Canada/mcmaster.ca/rdc/RDCGuide.pdf>

Manitoba Centre for Health Policy (est. 1991)

- A research unit in the University of Manitoba's Faculty of Medicine, which provides accurate and timely information to healthcare decision-makers, analysts and providers, offering services which are effective and efficient in maintaining and improving the health of Manitobans
- MCHP maintains and stewards population health research data stores, a robust collection of administrative, registry, survey, and other data primarily relating to residents of Manitoba. This includes administrative health data for virtually all contacts with the provincial health care system and the Manitoba Health Services Insurance Plan (including physicians, hospitals, personal care homes, home care, and pharmaceutical prescriptions)

Data Sources

- The Population Health Research Data Repository housed at MCHP combines Health, Education, Social, Justice and Registry data
- It is a comprehensive collection of administrative, registry, survey, and other data primarily relating to residents of Manitoba

Users

- Researchers, practitioners, programmers and students

Services

- The MCHP has online research tutorials for database and research skills.
- MCHP provides various courses to support health-services and population-based research
- MCHP offers knowledge translation in Manitoba by sharing evidence from published MCHP reports with health and social policy planners and decision makers throughout the province

Governance

- MCHP acts as a steward for the information housed in the Population Health Research Data Repository for agencies such as Manitoba Health
- MCHP is listed as a prescribed health research organization under Manitoba's Personal Health Information Act and Regulation

Access and Approval Process

- To gain access, Researchers must complete an MCHP Project Feasibility and Data Access Quote Request form and then submit a research proposal
- Researchers must also complete a yearly Accreditation session which provides a consistent overview of MCHP, and the data access and use process

Security and Privacy

- To maintain the privacy of Manitobans, all data provided by Manitoba Health is de-identified before being transferred to the Repository. Other privacy-enhancing technological safeguards include: firewalls, complex passwords, file encryption and restricted access protocols for the Repository
- Use of the data is regulated under the Manitoba Personal Health Information Act (PHIA) and the Freedom of Information and Protection of Privacy Act (FIPPA)

Reference:

Manitoba Centre for Health Policy (MCHP):

http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/

MCHP Privacy Code: http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/media_room/media/MCHP_privacy_code.pdf

http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/media/Story_of_MCHP.pdf



Institute for Clinical Evaluative Sciences (est. 1992)

- In Ontario, the Institute for Clinical Evaluative Sciences (ICES) is an independent, non-profit organization, whose core mandate is to conduct research that contributes to the effectiveness, quality, equity and efficiency of the health care sector.
- There are six ICES facilities currently located across the province (one central site in Toronto and five located at five different universities). One site in northern Ontario is currently under development.
- The ICES Data Repository is a collection of individual-level, de-identified and linkable data sets that encompasses much of the publicly-funded Ontario Ministry of Health and Long-Term Care (MOHLTC) services. Records for the Ontario population (eligible for universal health coverage) date back more than 20 years
- “ICES at Home” is a Secure Access Portal for Researchers to access data remotely

Data Sources

- ICES includes predominantly health sector data (including administrative, registry, vital statistics, demographic, survey, clinical and chart abstraction data). ICES has recently expanded into Children and Youth Services, Education, and Community Safety and Correctional Services

Users

- Historically, only ICES scientists and their designates (e.g., students) could access ICES data sets
- Recently, ICES launched “ICES Data & Analytic Services” to provide “research-ready linked health administrative data and analytic tools through secure virtual desktop infrastructure”
- The initiative is a component of Ontario’s SPOR SUPPORT Unit, and is intended to expand the user base to non-ICES researchers, students, policy makers and knowledge users.

Services

- Support and advice on health administrative dataset creation
- Delivery of reports and data analysis services
- Remote, online access to datasets and tools provided

Governance

- Governance is delivered through a voluntary Board of Directors, with their primary responsibilities including the definition of strategic direction and research priorities
- A Scientific Advisory Board advises the Board of Directors on quality, direction, scope and focus of work
- A Privacy Office addresses legislative and regulatory requirements
- A Program Leaders Committee reviews and votes on ICES scientist submissions.

Access and Approval Process

- Access can be granted by becoming an ICES scientist or by collaborating with an ICES scientist;
- Access is restricted on a project-by-project basis; Scientists must apply for and justify each element of information
- For the new Data & Analytic Services capability, ICES staff provide end-to-end support for all applicants, from application initiation to project closeout

Security and Privacy

- ICES implements privacy policies and practices required by the Information and Privacy Commissioner of Ontario. Under Section 45 of PHIPPA. These policies include implementation of physical and logical controls to control access to data (e.g., use of secure zones within ICES facilities, complex passwords and encryption)
- ICES has also adopted several principles for data protection related to data collection, data access, user training and maintenance of confidentiality (through anonymization)
- All data is de-identified promptly after it is collected.

Reference:

Institute for Clinical Evaluative Sciences

<http://www.ices.on.ca/>

Applied Health Research Question (AHRQ) 2013/14 Guidelines

<http://www.ices.on.ca/~media/Files/AHRQ/AHRQ-Guidelines-2013-14-21May13.ashx>

UK Administrative Data Research Network (in development)

- The ADRN is an initiative aimed at improving access to linkable administrative data from all sectors within the United Kingdom (UK) public service. Data will be housed within four Administrative Data Research Centres (ADRCs) (i.e., one in each of the four nations of the UK)
- The initiative has created a formal Administrative Data Service which will be the single point of contact for data providers and users. It includes a UK-wide Governing Board, a training and accreditation program, guidance on safe settings, public engagement strategies and partnerships with national statistical agencies
- The ADRN initiative is underway and requests for proposals have been issued for organizations interested in fulfilling the role of the AD Service and for organizations interested in hosting one of the four data centres
- Initially, its focus will be on improving access for academic and government researchers. However, it is intended to be a scalable solution and could in the future be extended to include other types of data and additional users

Data Sources

- Administrative data on education, health, employment, incomes, wealth, housing conditions, health incomes, benefits, parental employment, social background, childcare, etc

Users

- Academic researchers, government officials

Services

- The AD Service will provide a single point of entry for data providers and data users
- Facilitate the data acquisition process with providers
- Facilitate/coordinate data access requests by data users, and validate all requests to ensure only appropriate applications are considered
- Maintain list of accredited researchers and oversee accreditation process
- Training, covering topics such as: best practices for working in a secure data environment; legal and ethical issues related to use of individual records; methods of disclosure control, etc.

Governance

- A UK Governing Board will be established that includes directors of the ADRCs, funders and international experts in the field of administrative data use, senior government representatives from various departments, agencies, and data providers, and one lay member
- The Board will promote and facilitate safe research on administrative data for public and policy benefit. It will work with data providers to approve access requests and work to encourage collaboration across the ADRCs and other research data centres. Members will provide expertise on areas such as ethical standards, international experience, linkage methodology, safe setting security, and legal and other relevant issues.

Access and Approval Process

- Researchers must complete an accreditation process prior to gaining access
- Proposals for access to and linkage between administrative datasets are reviewed to validate they are legal, viable, and technically feasible and that the research skills are available
- Three levels of access to data are provided:
 - Secure ADRC facility, where access will be supervised by the ADRC support team
 - Researchers can submit statistical syntax to the ADRC support team who will run it on behalf of the researcher
 - Remote secure data access facilities which allow virtual access to datasets held in the ADRCs. No data would be transferred to these remote safe sites.

Security and Privacy

- Data provided through the ADRN will be anonymized by a third party prior to being received at one of the four centres
- Security is to be addressed through physical and logical security controls applied to each of the three access mechanisms described.

Reference:

Administrative Data Research Network:
http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf
 Administrative Data Taskforce - Structuring the Administrative Data Research Network:
http://www.esrc.ac.uk/_images/ADT_Technical_Group_Report_tcm8-26530.pdf

UK Health and Social Care Information Centre (est. 2013)

- The Health and Social Care Information Centre (HSCIC) was set up as an Executive Non Departmental Public Body (ENDPB) in April 2013 to act as a hub to collect, analyze and present anonymized and confidential national health and social care data on a secure online platform, as well as:
 - Implement and manage national IT systems for transferring, collecting and analyzing information
 - Publish rules to set out how the personal confidential information of patients should be handled and managed by health care staff and organizations
 - Build a library of ‘indicators’ that can be used to measure the quality of health and care services provided to the public
 - Act to reduce the paperwork doctors, nurses and care workers have to complete by ensuring that only essential data is collected
 - Help health and care organizations improve the quality of the data they collect and send by setting standards and guidelines
 - Create a register of all the information collected, produced and published in a range of different formats
- The HSCIC is also pursuing the creation of the “care.data” database which will integrate a much broader range of clinical and administrative data to be made available (in de-identified form) to citizens, patients, researchers, health system administrators and the private sector

Data Sources

- Administrative and clinical health data
- Public health data
- Social care data
- Workforce statistics

Users

- Academic researchers, healthcare commissioners and providers, volunteer organization, information intermediaries and commercial organizations including life science companies

Services

- HSCIC has a range of services to help data users access, link and create custom data in bespoke formats. Additional services include:
- Publication of health-related indicators, covering topics such as population health, quality, prescribing, social care and mental health.
- Publication of official statistics
- As-requested research reports
- Delivery of clinical audits (and publishing of results online)
- Publication of surveys (patients and staff)
- Implementation/operation of IT infrastructure.

Governance

- The Data Access Advisory Group (DAAG) is an independent group hosted by the Health and Social Care Information Centre which considers applications for sensitive or identifiable data

- The Health and Social Care (HSC) Act sets out the powers of the Secretary of State and NHS England to direct the HSCIC to carry out additional functions concerning:
 - The information functions of any health or social care body
 - Systems delivery functions.

Access and Approval Process

- Packaged reports containing aggregate data can be downloaded via the HSCIC website
- There is a formal approval process for data extracts with de-identified, person-level data (data that has been through a process of pseudonymisation but could still be re-identifiable)
- Approval times will vary depending upon the complexity of the request

Security and Privacy

- HSCIC is empowered by the Health and Social Care Act to collect and hold patient confidential information across the Health and Social Care system
- The security practices of the HSCIC are guided by a series of principles and policies
- A formal Privacy Impact Assessment has been completed and provides an overview of many of the security and information management controls in place

Reference:

Health and Social Care Information Centre (HSCIC): <http://www.hscic.gov.uk/>

HSCIC Business Plan: http://www.hscic.gov.uk/media/11860/HSCIC-business-plan-2013-14/pdf/80305_HSCIC_Business_plan_V1.0.pdf

HSCIC Privacy Impact Assessment: http://www.hscic.gov.uk/media/12931/Privacy-Impact-Assessment/pdf/privacy_impact_assessment_2013.pdf

UK Clinical Practice Research Datalink (est. 2012)

- The Clinical Practice Research Datalink (CPRD) is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA) and hosts NHS data and research services. CPRD services include linking anonymized NHS clinical data to aid in observational research and produce research outputs that translate into improvements in public health
- CPRD services include the following:
 - Data Services: Access request approval process, data linkage, data validation and real time data access
 - Interventional research services: Investigation of research proposal feasibility, recruitment at a site and patient level, electronic-data services, clinical trial linkage services, full clinical trial input, biological sample collection and linkage, and patient reported outcomes collection and linkage
 - Research services: Expert support staff to advise on research methodology, research governance and feasibility of research proposals, consultancy services including the provision of patient and/or healthcare professional input, and the ability to undertake research studies on behalf of customers

Data Sources

- Data collected by the CPRD includes; demographics, medical symptoms, signs and diagnoses, therapy, treatment outcomes, events leading to withdrawal of a drug or treatment, referrals to hospitals or specialists, laboratory tests, pathology results, lifestyle factors, socioeconomic status, patient registration, practice and consultation details
- All CPRD datasets contain coded data. The CPRD supply the necessary coding dictionaries and/or links to dictionaries to develop code sets and understand the variables used.

Users

- Academic, pharmaceutical, biotechnology and medical device companies and researchers

Services

- See summary above.

Governance

- The CPRD is hosted and managed by the Medicines and Healthcare products Regulatory Agency (MHRA)
- The Standard Operating Procedures (SOPs) which apply to CPRD dictate exactly how each stage in the process of making data available for research must be undertaken. SOPs are reviewed on a regular basis and compliance is mandatory and continuously monitored

Access and Approval Process

- Researchers can access primary care data online via CPRD GOLD. An extract tool enables, as specified, cuts of the data against a cohort or control group
- Researchers who have access to GOLD can also request a data cut from the GP SILVER dataset against a specific cohort query
- Every Research user is under a legal agreement which states the responsibility placed upon the research organization and their researchers for appropriate use of the data

Security and Privacy

- Personal identifiers of the datasets are not collected. CPRD uses methods of anonymization at more than one point in the process when making data available for a research project.
- UK and European law controls all access and use of data. Researchers can only access key datasets via standard security systems including a passphrase which changes every 30 seconds. CPRD requires a certificate on the computers of authorized researchers
- Physical security is ensured through a tier 3 data centre. Security measures are regularly reviewed and audited

Reference:

UK Clinical Practice Research Datalink : <http://www.cprd.com/>



AU Population Health Research Network (est. 2006)

- The Population Health Research Network (PHRN) is a nationwide data linkage infrastructure that supports secure and safe management and linkage of health information from around Australia
- The PHRN is a conglomeration of participant organizations that work within a defined governance framework of roles and responsibilities that align with federal and state privacy legislation and policies
- The PHRN network has the following components:
 - Program Office located in Perth, Western Australia
 - A network of Data Linkage Units (DLUs) that support each State and Territory
 - Two national Data Linkage units (which perform State-State, Commonwealth-Commonwealth and Commonwealth-State linkages)
 - Remote Access Laboratory (Secure Unified Research Environment) at the Sax Institute in New South Wales

Data Sources

- Data sources vary by State/Territory, but include:
 - Administrative and clinical health data
 - Vital statistics data
 - Education data
 - Social services data

Users

- Academic researchers

Services

- Data linkage services
- Provision of a secure research environment (SURE)
- Training programs for SURE users
- Secure data transfer services (SUFEX)
- Within each DLU, client services teams provide expertise on topics including design and cohort selection, completion of data applications, negotiation of contracts with data custodians and in some instances, archiving of data extracts at the completion of the study

Governance

- The PHRN Access Committee advises the Management Council on policies and processes regarding the issues of user eligibility, accreditation and conditional access
- The Program Office in Perth acts as the administrative hub and client support base, and coordinates development of the national data linkage infrastructure
- An Advisory Group provides advice to the main governing body around ethics, privacy and consumer/community issues

Access and Approval Process

- Applications are reviewed by a data custodian and a Human Research Ethics Committee to ensure the minimum necessary information is released
- Once approved, researchers sign a confidentiality agreement and are provided a project-specific unique Project ID
- The DLUs provide a linkage key to all data custodians for the approved dataset(s). Custodians then remove all identifiers, replace them with the linkage key and send the encrypted, de-identified data to the researcher
- Researchers can also access data remotely through the Secure Unified Research Environment (SURE) which has been developed by the Sax Institute (hardware physically located at, and controlled by, the SURE)

Security and Privacy

- Physical, IT, personnel and administrative privacy and security-enhancing policies and procedures are implemented (including encryption, firewalls, complex passwords, secure data transfers, and closed networks)
- The Secure Unified Research Environment (SURE) is a secure remote-access data research laboratory purpose-built for the analysis of routinely collected linked data
- SURE minimizes the risk of security breaches when conducting research by supplying a remote-access computing environment with comprehensive security features which replaces a researcher's local computing environment. Security features include a central server accessible only via an encrypted Virtual Private Network (VPN) through a firewall, and four means of researcher identification for access
- Personal information and content information is kept separate, with no party except the initial providing organization having access to the full dataset

Reference:

Population Health Research Network : <http://www.phrn.org.au/>



Mount Sinai Hospital (est. 1852)

- Mount Sinai Hospital, founded in 1852, is one of the oldest and largest teaching hospitals in the United States. The hospital's partnership network encompasses the Icahn School of Medicine at Mount Sinai, seven hospital campuses in the New York metropolitan area, and a large ambulatory footprint. The Health System is well established as a world leader for its excellence in research, patient care, and education across a range of specialties
- Significant investments were made in 2013 in the Icahn School of Medicine to provide advanced analytic technology through its Health System network, along with data scientists to amalgamate patient data into one central data warehouse that is used to develop predictive, multidimensional and risk models for the improvement of patient healthcare as well as analysis to support research at its dedicated school of medicine

Data Sources

- Detailed inpatient and outpatient data are extracted from transactional systems. The data warehouse is sourced from 20 transactional systems, encompassing over three million patients & approaching one billion facts

Users

- Patients, Researchers, Practitioners

Services

- Resources offered include:
 - Biostatistics, Epidemiology, and Research Design (BERD) Program offers educational, statistical advisory and consultative services, for study design, biostatistics, and bioethics
 - The International Center for Health Outcomes and Innovation Research (InCHOIR) offers expertise in the design, conduct, and analysis of randomized trials of novel interventions
 - The Centre for Biomedical Informatics (CBI) program focuses on translating research discoveries into clinical practice
- A web-based learning module educates researchers regarding compliance to protocols and data management requirements

Governance

- Partnership with the Icahn School of Medicine
- All studies require review and approval of the Research Ethics Board

Access and Approval Process

- The Mount Sinai Hospital Research Ethics Board is responsible for approving ~ 300 new studies each year involving human subjects and overseeing all such research under the auspices of Mount Sinai Hospital Board
- Access is granted through a formal application process. Expedited reviews can take approximately three to four weeks for Mount Sinai Researchers. For REB-approved research, datamarts tailored to specific eligibility and data analysis requirements are prepared. Datamarts may be de-identified, or may contain PHI with appropriate approvals

Security and Privacy

- The Mount Sinai Data Warehouse (MSDW) includes clinical, operational, and financial data derived from patient care processes and is compliant with HIPAA and New York State privacy and security regulations
- Additional security at Mount Sinai includes the following:
 - Portal security is Role-Based Access Control based on industry standards
 - Portal approved by HIPAA & Mount Sinai IT Security
 - Secured data transfer via HTTPS protocol
 - Servers are located in the Mount Sinai Data Center
 - Database access requires user and role authorization
 - All access (views, edits, deletes) to data within the data warehouse is logged and monitored.

Reference:

Mount Sinai Hospital: <http://www.mountsinai.org/>

Icahn School of Medicine - Research Information Technology:

<https://icahn.mssm.edu/about-us/services-and-resources/computer-services/resources/research-information-technology>

- Intermountain Healthcare is a nonprofit health system with a broad range of clinics and services, health insurance plans from SelectHealth, 22 hospitals, and over 800 physicians in the Intermountain Medical Group. Intermountain has over 33,000 employees creating standardized data to facilitate a community of medical researchers
- Intermountain has an extremely comprehensive data repository of clinical and financial data from its 22 hospitals and 185 clinics, covering a 40-year span with more than two trillion unique medical data elements

Data Sources

- Intermountain offers comprehensive access to clinical and administrative data
- Through a partnership with Deloitte, Intermountain has developed OutcomesMiner, which analyzes the relationships between healthcare conditions and outcomes for a patient, and PopulationMiner that draws on Intermountain's repository of clinical, financial and operational data stored in a warehouse that supports patient-outcome analysis

Users

- Researchers, Practitioners and Patients

Services

- Intermountain offers various training courses for Researchers, that cover executive level training for tools to implement quality improvement programs, quality improvement aimed at local clinicians, and courses on tools and principals of quality improvement and research for Intermountain employees
- Clinical Studies Managers provide expertise and support for researchers in managing and coordinating clinical studies
- Intermountain also offers a Statistical Data Centre (SDC) which consists of trained statisticians who are available to assist researchers

Governance

- Institutional Review Boards (IRBs) review and monitor all research conducted at Intermountain Healthcare

Access and Approval Process

- Intermountain lists specific and direct contacts where users can submit a data request.
- All submissions are reviewed by the IRB. IRB submissions are accepted online through the Research Management System. Users are provided with an Intermountain LDAP login and password to access their application internally and externally

Security and Privacy

- The Institutional Review Board (IRB) at Intermountain reviews proposed research protocols to ensure that security and privacy principles are consistently applied. The IRB is responsible for the following:
 - Ongoing review of approved protocols
 - Ensuring the ethical conduct of research involving human subjects
 - Review and approval of human subject applications
 - Monitoring of adverse events for research including human subjects

Reference:

Intermountain Healthcare: <http://intermountainhealthcare.org/>

Intermountain Generating Population Health Data Revenue:

<http://www.healthleadersmedia.com/page-1/TEC-297956/Intermountain-Generating-Population-Health-Data-Revenue##>

Appendix C - Phase 1 Activities Supporting the Creation of the BC CDI

The Ministry of Health has a long-term vision for increasing the strength of data-informed health sector management by enhancing the integration of health information across the sector. This vision aligns with the vision for the BC CDI and similarly will need to be implemented in phases over several years. It includes:

- an initial focus on improving access to consolidated administrative data for research purposes;
- greater coordination and use of primary data that can be linked to administrative data;
- enhanced coordination in the safe collection, storage and use of researcher data; and
- expansion, over time, to include data from health authorities and medical records from health care practitioners

Population Data BC is best positioned to support the Ministry of Health and, by extension the Province, in implementing this first phase of work. Population Data BC has a well-established relationship with the Ministry of Health and currently plays a significant role in the provision of health data for research purposes in BC. It is a multi-university data and education resource that facilitates research on the determinants of human health, well-being and development. The organization facilitates researcher data access requests, provides access to linked, anonymized data sets to approved researchers in a physical setting at the University of British Columbia (UBC) and remotely via a “Secure Research Environment”. It supports users by providing a range of services and analytical tools to obtain approved access to relevant datasets, analyse these datasets and interpret results. Population Data BC also offers a range of services to data providers. The organization works with providers to establish appropriate Data Sharing Agreements and then supports data access by facilitating the data request process, providing an environment for the secure storage, anonymization and linkage of data and providing users with physical and virtual access to these data sets in a secure manner.

To implement Phase 1 of the Ministry of Health’s data integration vision and, by extension, Phase 1 of the BC CDI vision, the following activities are recommended as next steps:

- **Enhance security and privacy capabilities within HealthIdeas to support secure remote access**
Complete security enhancements to HealthIdeas to enable access by Population Data BC. This component includes:
 - Implementation of a “Secure Access Environment” with the Ministry of Health that enables internal users to conduct analysis in a protected and monitored environment. Users will not be able to remove data from this environment unless reviewed and approved by the Ministry.
 - Implementation of a capability to detect inappropriate activity in real time
- **Enable access to the Ministry of Health’s HealthIdeas data warehouse by Population Data BC**
Currently, Population Data BC receives extracts from HealthIdeas on a periodic basis. As a result, the data provided to researchers is not as current as the data available within the Ministry’s Data Warehouse. To provide the most current data possible to approved researchers, it is proposed that direct access for select Population Data BC analysts be provided.

This initiative was designed to provide short-term benefits to the research community by enabling access to more current health data and by providing Population Data BC with access to a broader range of data maintained within HealthIdeas. This project does not introduce any new collection, use or disclosure of personal information or any new information. As such, the existing Privacy Impact Assessments and Information Sharing Agreements currently in place will continue to apply and govern access to this information.

