THE WINSTON CHURCHILL MEMORIAL TRUST OF AUSTRALIA

Report by Dr Felicity Flack
2011 Churchill Fellow

Data Linkage Research: Maximising Public Benefit and Minimising Risks to Individual Privacy

I understand that the Churchill Trust may publish this report, either in hard copy or on the internet or both, and consent to such publication.

I indemnify the Churchill Trust against any loss, costs or damages it may suffer arising out of any claim or proceedings made against the Trust in respect for arising out of the publication of any Report submitted to the Trust and which the Trust places on a website for access over the internet.

I also warrant that my Final Report is original and does not infringe the copyright of any person, or contain anything which is, or the incorporation of which into the Final Report is actionable for defamation, a breach of any privacy law or obligation, breach of confidence, contempt of court, passing-off or contravention of any other private right or of any law.

Signed  

Dated 11.12.2011
Contents
Introduction .............................................................................................................. 1
Executive Summary .................................................................................................. 2
Programme ............................................................................................................... 3
Aims and Objectives ............................................................................................... 3
Main Body ............................................................................................................... 4
  Information Services Division National Health Service Scotland .................. 6
  Scottish Health Informatics Programme ....................................................... 7
  Health Information Research Unit, University of Swansea ......................... 9
  Ontario Agency for Health Protection and Promotion ............................... 11
  Institute for Clinical Evaluative Sciences ..................................................... 12
  Manitoba Centre for Health Policy ............................................................... 14
  Population Data BC ......................................................................................... 17
Conclusions and Recommendations ............................................................... 20
Glossary .................................................................................................................. 22
Introduction
My professional career has always revolved around health and medical research. Initially as a molecular microbiologist but for the last ten years I have worked as a research administrator with a particular interest in research governance and ethics. Two and half years ago I took the opportunity to work for the Population Health Research Network, a new national initiative to build data linkage infrastructure for all the States and Territories in Australia.

The Population Health Research Network is a unique initiative which will result in better use of Australia’s data resources and improvements in health and human services. As the Population Health Research Network is a new enterprise it is possible to establish the best possible techniques and processes for minimising privacy risks and maximising public benefit for research using linked data across Australia. Receiving the Churchill Fellowship has enabled me to visit and learn from the best linkage organisations in the world. I intend to use everything that I have learned and the contacts that I have made to contribute to developing the Population Health Research Network.

It has been a privilege and a pleasure to travel and represent the Winston Churchill Trust and I am very grateful to the Winston Churchill Trust for giving me the opportunity.

I would also like to thank my two referees, Dr Merran Smith for introducing me to the world of data linkage and guiding me through its complexities and Professor Fiona Stanley for being an inspiration and a support throughout my career. Finally many thanks to all the amazing people from data linkage organisations in Scotland, Wales and Canada who opened their doors, generously gave of their time and shared their passion and experience.
Executive Summary

Felicity Flack
105 Hay Street
Subiaco WA 6008
Manager, Policy and Client Services
Population Health Research Network
08 6389 7302

Data Linkage Research: Maximising Public Benefit and Minimising Risks to Individual Privacy

Highlights
- Attending the Scottish Health Informatics Programme Conference, St Andrews Scotland;
- Attending a meeting with Scottish Government officials in the Dome Room at General Register House in central Edinburgh;
- Visiting the Health Informatics Research Unit at Swansea University;
- Visiting all the major data linkage units in Canada.

Recommendations
- That the Population Health Research Network considers ways to minimise the amount of time and paperwork involved in the application and approval processes for access to linked data.
- That the Population Health Research Network considers ways to streamline the data extraction process. Options to consider include data repository-type solutions as well as funding for dedicated staff for the data custodians.
- That the Population Health Research Network mandate privacy and ethics training for all researchers accessing linked data through the Population Health Research Network infrastructure
- That all Population Health Research Network participants consider establishing the role of information governance or privacy officer if the position does not already exist.

Implementation and Dissemination
- Present my experiences and findings to relevant Population Health Research Network committees including the Operations Committee and Management Council;
- Make my report available on the Population Health Research Network website;
- Use the information gathered about client services and minimising barriers to accessing linked data to assist in the design and development of the Population Health Research Network cross jurisdictional application process and other client services;
- Use the information gathered about metadata and data documentation to contribute to the Population Health Research Network Metadata Framework;
- Discuss my experiences and findings with colleagues.
Programme
7 - 12 September 2011
Scotland
- Scottish Government (Education and Health) and General Register Office for Scotland (Census)
- Information Services Division NHS Scotland
- Exploiting Existing Data for Health Research, International Conference, University of St Andrews Scotland
- Health Informatics Centre, Dundee University

13 - 15 September 2011
Wales
- Health Information Research Unit, University of Swansea

19 - 29 September
Canada
- Ontario Agency for Health Protection and Promotion
- Institute for Clinical and Evaluative Sciences
- Manitoba Ministry of Health
- Manitoba Centre for Health Policy
- Population Data BC

Aims and Objectives
The aims of this Fellowship are to:
- Increase my understanding of different types of linkage systems;
- Increase my understanding of different governance processes used by these linkage systems;
- Understand how the choice of linkage system and/or governance processes can maximise public good and minimise risks to individual privacy.

The objectives of this Fellowship are to:
- Document the governance processes used in a range of different linkage systems;
- Identify which characteristics of a linkage system and/or associated governance processes that maximise the use of linked data for public good;
- Identify which characteristics of a linkage system and/or associated governance processes that minimise the risks to individual privacy.
Main Body

What is Data Linkage?
Information about our health, education and welfare is recorded throughout our lives as we come in contact with a number of agencies, including hospitals (public and private), health departments, schools and other government departments. The collection of this data is often required under legislation and the information is stored in secure computer databases within the agencies.

Data linkage is a technique for creating links between pieces of information that are thought to relate to the same person, family, place or event. This function is often performed by specialist data linkage units which facilitate access to linked data to enable research for the public benefit.

Research using linked data has been regularly used in the health sector to provide us with a picture of the health and well-being of our community. The results from these studies have provided valuable information on the causes of and risk factors for disease, the evaluation of new approaches to prevention, early intervention and treatment, and the assessment of novel ways of delivering health services. More recently, the linkage of data from areas including education, criminal justice and child protection has been undertaken to provide insight into other factors that can influence an individual’s overall welfare.

Research using linked data of this kind can be of enormous benefit to our community but it does require the use of information, often sensitive information about people’s health without their consent. Therefore it is essential that organisations that undertake data linkage should make every effort to minimise the risks to people’s privacy and maximise the benefits to the Australian people.

The Population Health Research Network
The Australian Government Department of Innovation, Industry, Science and Research have recently funded a new national initiative to build the capability for research using linked data in all States and Territories. The initiative is called the Population Health Research Network. This means that there is a unique opportunity to establish the best possible techniques and processes for minimising privacy risks and maximising public benefit for research using linked data across Australia.

International Data Linkage Groups and Organisations
By their nature, data linkage systems deal with large volumes of data and require complex organisational and technical infrastructure. Bringing together information from different sources often requires many different organisations to collaborate to develop and maintain a data linkage system. There are a range of examples of data linkage systems around the world. The structures of these systems reflect their purpose and the organisational environment within which they were developed.
There are four main functions that data linkage organisations can perform. They are:

1. Storage and maintenance of data collections
2. Data linkage
3. Facilitation of access to linked data
4. Analysis and evaluation of linked data

1. **Storage and maintenance of data collections**
   Data linkage cannot occur unless there is data collected, stored and maintained which is made available for linkage. The collection of data is independent of data linkage and occurs whether or not linkage occurs.

   In this function data collections are stored and maintained. Data collections can be stored:
   - Individually by the data custodian responsible;
   - In a repository, where multiple collections are stored and maintained in a central place.

2. **Data linkage**
   In this function links are created between pieces of information that are thought to relate to the same person family place or event.

   Access to complete records is not required to create links between pieces of information. Usually a small number of variables such as name, address, sex and date of birth are sufficient.

3. **Facilitation of access to linked data**
   In this function a range of tools and services are available to researchers to facilitate access to linked data for research and analysis. Examples of tools and services include:
   - Information about data collections and their availability for linkage (metadata);
   - Assistance with research design and implementation;
   - Coordination of approval processes for access to data;
   - Extraction and merging of datasets for approved research projects;
   - Technology for the secure transfer, storage and management of research datasets (e.g. remote access laboratories);
   - Training for researchers e.g. privacy training.

4. **Analysis and evaluation of linked data**
   In this function linked data are analysed and evaluated to increase knowledge with the ultimate goal that improved knowledge will result in improved outcomes and services in areas such as health, education and social well-being.

Organisations involved in data linkage may perform one or more of these functions. One of the main drivers behind which functions an organisation performs and how they are performed is the protection of privacy and confidentiality.
A range of different data linkage systems are in operation in the United Kingdom, Canada and Australia. The aim of this Fellowship is to compare and contrast these different systems.

Information Services Division National Health Service Scotland

Introduction
http://www.isdscotland.org/index.asp

“The Information Services Division (ISD) is a division of National Services Scotland, part of NHS Scotland. ISD provides health information, health intelligence, statistical services and advice that support the NHS in progressing quality improvement in health and care and facilitates robust planning and decision making.”

The Scottish Health Information Service (SHIS) and the Scottish Health Information Service for Research (SHIS-R) are parts of ISD. The role of SHIS is to enable Scottish Health Boards to perform research and analysis using the NHS Corporate Data Warehouse and the ISD data marts. SHIS also has a Record Linkage Team which links nationally held data to other datasets.

Storage and Maintenance of Data Collections
A wide range of identifiable health data are brought together and stored in the National Database at ISD. The data collections include hospital morbidity data, prescribing data, allied health data, cancer registry data and many others.

Data marts are a new initiative of SHIS. A data mart is a collection of tables held in the data warehouse that satisfy the reporting needs of a business area e.g. the ACaDMe Data Mart contains information about acute cancer deaths and mental health.

Data Linkage
ISD creates and holds two main linked datasets, the “Acute Dataset” and the “Maternity Dataset”. These are regularly linked with other datasets such as the Scottish Health Survey and the Scottish Longitudinal Study. In addition they also have the capability to link datasets together from outside ISD e.g. research datasets.

Facilitation of access to linked data
ISD facilitates access to linked data for researchers and analysts in government agencies and universities. This is achieved by the provision of a range of information and services which include:

- Website – provides information about the data available and the application and approval processes;
- Coordination of applications and approvals;
- Metadata – including a data dictionary, data manual and validation manual.
Analysis and evaluation of linked data
The Healthcare Information and Epidemiology and Statistics Groups at ISD have a primary role in the analysis and dissemination of linked data.

Scottish Health Informatics Programme

Introduction
http://www.isdscotland.org/index.asp

The Scottish Health Informatics Programme (SHIP) is a Scotland-wide research platform for the collation, management, dissemination and analysis of anonymised electronic patient records. It is a collaboration between NHS Scotland and a number of academic partners. It is also an interdisciplinary collaboration which covers areas including health research, privacy and confidentiality, law and public engagement. SHIP is funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council and aims to:

- Provide access to a new national research facility, firmly embedded within and supported by NHS Scotland, providing the basis for numerous future studies using electronic patient records;
- Create a research portal for electronic patient records already held by NHS Scotland that will provide rapid, secure, access to the type of data that clinical scientists require;
- Develop and evaluate systems that work across institutional boundaries to allow linkage between large, federated, third party research datasets and the NHS research portal.

Storage and Maintenance of Data Collections
SHIP does not store or maintain data collections

Data Linkage
The proposed linkage model for SHIP is that once a researcher has gained all the necessary approvals for a data linkage project, the datasets will be sent to an indexing service responsible for adding a variable to records within the datasets so that each individual could be uniquely identified and the datasets anonymised. The Community Health Index number will be used as an indexing variable.

Facilitation of access to linked data
One of the primary roles of SHIP is to facilitate access to electronic patient records for researchers. This is achieved by the provision of a range of information and services which include:

- Website – the SHIP Information Governance Toolkit is still being developed but will assist researchers to navigate the complex legal and ethical requirements which govern access to personal information for research;
- Coordination of applications and approvals
The SHIP Research Coordinator will assist researchers through the necessary approval processes;

- SHIP Safe Haven – a physical or virtual place where sensitive data can be analysed and risk of disclosure is reduced by controlling who can have access, what data they can analyse and what outputs can be taken away;
- Training – a programme of training workshops and seminars is offered to researchers as well as a biennial conference.

**Analysis and evaluation of linked data**
In the SHIP model research, analysis and evaluation is carried out by researchers either in their home institution or within a designated safe haven.

**Information Governance**
The SHIP Information Governance Working Group has developed a document called "SHIP Guiding Principles and Best Practices" which is designed to be used as a template for good governance. It covers the following topics:

- Public Interest
- Privacy
- Consent
- Anonymisation
- Authorising/advisory bodies
- Governance
- Access
- Trusted Third Parties
- Data Controllers and Data Processors
- Clinical Trials
- Cross-sector sharing
- Data sharing agreements
- Public and stakeholder engagement
- Sanctions
- Benefit Sharing

**Consumer and Community Participation**
One of the core SHIP Programmes is around public engagement. The aims of this programme are:

- To synthesise existing evidence on citizens' attitudes towards and willingness to share personal data for research;
- To generate new evidence on the acceptability of different levels of data sharing under varying conditions;
- To engage the wider public with the aims, processes and outcomes of the SHIP through specific consultation exercises and workshops;
- To link the public engagement activities into the development of governance frameworks;
- To examine novel methods of making health data available to the public.
Health Information Research Unit, University of Swansea

Introduction
http://www.swansea.ac.uk/medicine/ils/healthinformationresearchunit/

The Health Information Research Unit (HIRU) was established by the College of Medicine, University of Swansea and is funded by the Welsh National Institute of Social Care and Health Research. The aim of HIRU is to realise the potential of electronically held, person based, routinely collected information for the purposes of conducting and supporting health related research.

HIRU is part of the University of Swansea and is therefore subject to the same oversight and governance as all University departments. It also has three additional layers of governance:

- HIRU Advisory Board – an independent Board comprised of leaders in the field of health information and research in Wales. The Board performs a strategic and advisory role, including ensuring that HIRU activities are meeting the needs of the Welsh community;
- HIRU Consumer Panel – comprised of service user and carer representation on strategic committees and research study steering groups. The Panel provides advice on HIRU activities from the perspective of service users/carers and advice about how best to engage with consumers and the community;
- HIRU Management Team – comprised of the HIRU Directors and senior staff. The Management Team manage the day to day operations of HIRU.

Storage and Maintenance of Data Collections
HIRU has established the Secure Anonymised Information Linkage (SAIL) system which has two main parts, a linkage function and an anonymised databank (data repository). The databank is housed at the University of Swansea in the College of Medicine’s IBM Supercomputer.

The SAIL Databank includes data from a wide range of sources. Examples of the types of data in the Databank are:

- Registers of births and deaths
- Hospital Episode statistics
- Breast screening data
- Cancer Registry
- General Practice primary care data

HIRU are continually working to expand the type and geographical coverage of the data in the Databank. A complete list of the data collections included can be found at http://www.swansea.ac.uk/ils/research/healthinformationresearchunit/.

Data Linkage
In the SAIL system the data linkage function is performed by the NHS Wales Informatics Service. The separation of the linkage function from the storage and maintenance of the anonymised databank is central to the protection of privacy and
confidentiality. This separation of functions ensures that the data providers are the only ones who have access to the complete, fully identified data collections and that the HIRU staff and the researchers that use the SAIL Databank do not have access to identifiable information.

Once a data provider has agreed to contribute to the SAIL system and signed the Data Collection/Confidentiality Agreement they sort the data into two files. The first file contains identifiable data such as name, address and date of birth. This file is sent using secure HTTPS upload facilities to the NHS Wales Informatics Service. The second file contains content information such as diagnosis and interventions. This file is sent by the same secure method to HIRU to be included in the SAIL Databank.

The NHS Wales Informatics Service use the identifiable information they receive to match it to NHS numbers in the NHS Administrative Register. They then replace the NHS number with a unique encrypted code called the Anonymous Linking Field (ALF). The ALFs are sent to HIRU where they are encrypted again to ensure that neither NHS Wales Informatics Service nor HIRU can decrypt the ALFs. HIRU can then use the double encrypted ALF to merge data, from different anonymous collections, about the same individual for research projects.

Facilitation of access to linked data
One of the primary roles of HIRU is to facilitate access to linked data for researchers. Access is facilitated by the provision of a range of information and services which include:
- Website – which provides researchers with information about what data is available through the SAIL Databank and how to apply for it;
- Coordination of applications and approvals;
- Assistance with project design;
- The SAIL Gateway – a remote analysis laboratory;
- Training – in analysis of linked data; use of the SAIL Gateway.

Analysis and evaluation of linked data
HIRU also conducts research using the SAIL system with the aim of improving service delivery. To date their research has covered a broad range of topics from emergency medicine to obstetrics and gynaecology. A list of publications is available on the HIRU website.

Information Governance
HIRU have a very robust information governance framework which includes:
- Legal and Regulatory Compliance
- Risk Assessment and Risk Management
- Audit, Reviews and Reporting Guidelines
- Incident and Breach Management
- Awareness, Education and Training

Consumer and Community Participation
HIRU have recently set up a Consumer Panel which has the following aims:
• Act as an advisory panel to HIRU on issues in research from the perspective of service users and carers;
• Provide HIRU with the service user/carer view on data protection issues;
• Advise on how best to engage with the public/service users and carers;
• Advise on how to recruit people to project steering groups;
• Discuss proposals for projects and receive updates on research findings.

Ontario Agency for Health Protection and Promotion

Introduction
http://www.oahpp.ca/index.html

The Ontario Agency for Health Protection and Promotion (OAHPP) is the public health agency for Ontario which sits at arm’s length from Government. It is responsible for providing scientific advice and practical tools to support health care providers, the public health system and partner ministries in making informed decisions and taking informed action to improve the health and security of the people of Ontario. It provides expert scientific and technical support relating to infection prevention and control; surveillance and epidemiology; health promotion, chronic disease and injury prevention; environmental and occupational health; health emergency preparedness; and public health laboratory services.

Storage and Maintenance of Data Collections
The OAHPP collects and stores a range of data that mostly relates to public health and disease surveillance. In particular they run the Public Health Ontario Laboratories which provide clinical and environmental testing throughout Ontario with a focus on infectious and communicable diseases such as tuberculosis, HIV, SARS, Legionnaires’, influenza and West Nile Virus infections. This is a very rich source of clinical information and biospecimens.

Data Linkage
The OAHPP is currently in the early stages of investigating the possibilities for linking the data they collect both within the organisation and also with data collections not held by OAHPP such as data collections from Ontario Ministry of Health.

Consumer and Community Participation
The OAHPP have planned an extensive program of community consultation in the development of their data linkage program.
Institute for Clinical Evaluative Sciences

Introduction
http://www.ices.on.ca/

The Institute for Clinical Evaluative Sciences (ICES) is an independent, non-profit corporation and registered charity, whose main purpose is to conduct research that contributes to the effectiveness, quality, equity and efficiency of health care and health services in Ontario. In order to do this it maintains a database of health and population data.

ICES is governed by a voluntary Board of Directors and is funded by the Ontario Ministry of Health and Long Term Care as well as receiving peer-reviewed grants from provincial and national organisations.

Storage and Maintenance of Data Collections
ICES houses a broad range of population-based data collections covering a number of areas including:

- Health services data
- Population and demographic data
- Disease cohort/registry data
- Care Provider, Facility, Geography, Coding and Other Data

A partial list of the data collections can be found at
http://www.ices.on.ca/webpage.cfm?site_id=1&org_id=26&morg_id=0&gsec_id=5314&item_id=5322

Data Linkage
Once ICES has signed a data sharing agreement with a data provider, the data provider sends the data in encrypted form to ICES through secure sockets layer virtual private network. The dedicated covenantees at ICES receive the dataset and unencrypt it. Most datasets include the Ontario Health Insurance Plan (OHIP) health card number. This unique number is used by each individual to receive health services in Ontario. If this number is part of the dataset, the data providers do not usually send the names and addresses. The covenantees use the OHIP number to link the data to the healthcare number list. The OHIP number is then replaced with a unique number called the ICES Key Number or IKN. The covenanteer’s specific role is to match the identifying information on the incoming data set to the ICES data, assign a common ICES key number and strip off the identifying information. They do not perform data analysis on the de-identified data.

Facilitation of access to linked data
ICES has developed a range of tools and services to assist researchers to access their linked data resources. These include:

- Website
  The ICES website includes information for researchers about the ICES data holdings, how to apply to become an ICES scientist or collaborate with an ICES scientist, privacy, education opportunities and events, publications and outcomes of ICES research projects;
- **Metadata**
  ICES has a comprehensive metadata repository that is used by ICES scientists and is regularly updated. New data may be de-identified and integrated with ICES data through data sharing agreements with data custodians (data providers);

- **Coordination of applications and approvals**
  ICES data can only be accessed by ICES scientists or by collaborating with an ICES scientist. There is an internal application process which involves a review of the proposal and the completion of a privacy impact assessment. Individual projects are not required to be reviewed by a Human Research Ethics Board. ICES reports all projects to a Human Research Ethics Board and a sample of the projects are reviewed;

- **Assistance with project design**
  The ICES scientists and data analysts have extensive experience in the analysis of linked data and the data collections held by ICES. New ICES scientists undertake a type of apprenticeship where they work closely with an ICES core scientist. A probationary period must be completed before scientists are granted full access to the data. All projects are also assigned an experienced analyst to work on the project;

- **Data Access**
  Data can be accessed at the ICES Sunnybrook campus or at one of the ICES satellite sites at Queen's University and the University of Ottawa. New expansion sites at the University of Western Ontario and the University of Toronto will open in 2012.
  A new initiative, the Ontario Cancer Data Linkage Project (CD-Link) is trialling the release of data outside of ICES. Project specific data extracts are put together and modified so that they can be released outside of ICES. Specifically, personal identifiers are removed, other identifiers (e.g., institution) are scrambled, and all dates more specific than year are converted to number of days relative to the index date. Some variables (e.g., age, geographic regions) may also be categorized;

- **Training**
  ICES offers a number of training and education opportunities for researchers including weekly research rounds, symposia and workshops.

**Analysis and evaluation of linked data**

The central role of ICES is to conduct research. This is clearly identified in their mission and goals.

"The Institute for Clinical Evaluative Sciences (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."

"In support of our mission, ICES' key objectives are to:

- Carry out population-based health services research that is relevant to clinical practice and health policy development;"
• Document province-wide patterns and trends in health care delivery;
• Identify major determinants of disease;
• Assess the effectiveness and safety of a wide range of healthcare interventions;
• Develop and share evidence to inform decision-making by policy makers, managers, clinicians, planners and consumers;
• Promote linkages among health services researchers and decision-makers;
• Train researchers and promote a wider understanding of clinical epidemiology and health services research.”

Information Governance
• Legal and Regulatory Compliance
  ICES operates in a complex legislative environment which includes both Provincial and Federal legislation and regulations. In particular, Ontario’s Personal Health Information Protection Act as well as the federal Privacy Act and the Personal Information Protection and Electronic Documents Act. The Chief Privacy Officer oversees compliance with the legislative framework;
• Risk Assessment and Risk Management;
  ICES has established a formal risk management framework, which specifically addresses risks to information. The framework has been implemented, and is enabled by a range of processes and tools, which include an ICES risk assessment tool and security risk register. Together these create a climate of continuous detection and mitigation of information risk.
• Audit, Reviews and Reporting Guidelines;
  ICES undergoes a site audit and review of all practices, policies and procedures by the Information and Privacy Commissioner of Ontario every three years. ICES also undergoes annual electronic security audits by external security experts to ensure the integrity of its systems;
• Incident and Breach Management
  Requirements for the management of incidents, breaches and complaints are outlined in policy and procedure documents as well as in data transfer and confidentiality agreements.

Manitoba Centre for Health Policy

Introduction
http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departments/units/mchp/

The Manitoba Centre for Health Policy (MCHP) is a research unit in the University of Manitoba's Faculty of Medicine. MCHP's purpose is to conduct research on the health of the people of Manitoba. In order to do this it maintains an de-identified database of health, education and human services data.
As part of the University of Manitoba it is subject to the same governance and oversight as other parts of the University. In addition it has an Advisory Board whose membership consists of the Director of MCHP, senior staff from the University of Manitoba, representatives from a range of government departments and researcher representatives; some from other leading academic institutions.

**Storage and Maintenance of Data Collections**

MCHP houses the Population Health Research Data Repository which is a comprehensive collection of administrative, registry, survey and other databases with information about the people of Manitoba. It covers a broad range of areas including:

- Health
- Education
- Social
- Justice

A complete list of the data collections included can be found at [http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/repository/index.html](http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/repository/index.html).

The data held in the Repository are de-identified i.e. they do not include names and addresses. The Repository is regularly updated and MCHP staff continually work with data custodians to add new collections.

**Data Linkage**

The data linkage function for the MCHP is performed by Manitoba Health. The Personal Health Identification Number (PHIN) is the unique identifier assigned to each individual registered to receive health care by the Ministry of Health. It is used to link records across the different datasets. Both probabilistic and deterministic linkage techniques are used. Clerical review is performed as necessary. Before the data is sent to the Repository at MCHP the PHIN numbers are scrambled and the names, addresses and other direct identifiers are removed. However, typically full data of birth, sex and postal code are retained on the data so stringent protocols for protecting the privacy of individuals in the data are maintained.

**Facilitation of access to linked data**

MCHP has developed a range of tools and services to assist researchers to access their linked data resources. These include:

- **Website**
  - The MCHP website is very comprehensive and includes information for researchers about data, application and approval processes, privacy, study design and publications and outcomes of MCHP research projects;
- **Metadata**
  - MCHP Repository staff systematically develop and expand the documentation about the data and the data collections. This includes a data quality framework as well as an extensive concept dictionary and glossary;
- **Coordination of applications and approvals**
MCHP staff help coordinate the application process to the Health Research Ethics Board, Health Information Privacy Committee and data custodians;

- Assistance with project design
  The in-house MCHP data analysts have extensive experience in the analysis of linked data and the data collections in the Repository. They provide researchers with advice about research design as well as assistance with the analysis of data;

- Data Access
  Researchers can only access the Repository on site at MCHP. However, a pilot of remote access sites is currently underway. This allows researchers to see a pre-prepared data extract specific for their project;

- Training
  MCHP offers a number of training and education opportunities for researchers including weekly researcher meetings, statistical software workshops, geographic information systems workshops and mandatory privacy training. The University of Manitoba offers a course entitled the Epidemiology of Health Care.

Analysis and evaluation of linked data
As described in its mission statement MCHP is "a research centre of excellence within the Department of Community Health Sciences, Faculty of Medicine at the University of Manitoba that conducts world class population-based research on health services, population and public health, and the social determinants of health". Research is an important role of the Centre. A detailed description of MCHP’s reports, publications and presentations can be found on their website.

Information Governance

- Legal and Regulatory Compliance
  MCHP operates in a complex legislative environment which includes both Provincial and Federal legislation and regulations. In particular, Manitoba’s Freedom of Information and Protection of Privacy Act and Personal Information Protection Act as well as the federal Privacy Act and the Personal Information Protection and Electronic Documents Act;

- Risk Assessment and Risk Management
  The Manitoba Ombudsman’s office has a risk assessment tool that MCHP completes on a regular basis;

- Audit, Reviews and Reporting Guidelines
  The Manitoba Ombudsman requires a privacy audit to be conducted every two years. Information Technology security audits are conducted on an ad hoc basis. The University of Manitoba has conducted internal audits, as well as the Provincial Audit Offices;

- Incident and Breach Management
  Manitoba legislation requires everyone dealing with personal health information to complete an organization level orientation that covers management of incidents, breaches and complaints. Requirements for the management of incidents, breaches and complaints are outlined in policy and
procedure documents as well as in data transfer and confidentiality agreements.

Knowledge Translation
An important part of MCHP’s mission is that “MCHP promotes a collaborative environment to create, disseminate and apply its research. The work of MCHP supports the development of policy, programs and services that maintain and improve the health of Manitobans”. This focus on knowledge translation is central to the work of MCHP.

There are several arms to the knowledge translation work that is done including:
- The close working relationship with Manitoba Health who contract MCHP to conduct six research projects each year;
- The Need to Know Team which works with regional health authorities to foster research of relevance to regional communities and agencies;
- Regular workshops with Manitoba Health program managers to support planning and uptake of research knowledge.

Population Data BC

Introduction
http://www.popdata.bc.ca/

Population Data BC is a multi-university, data linkage and education resource facilitating interdisciplinary research on the determinants of human health, well-being and development. It has three main university partners, the University of British Columbia, the University of Victoria and Simon Fraser University.

The aim of Population Data BC is to provide researchers with access to linkable data from a range of sources such as health, education, early childhood development, workplace and the environment and the training required to conduct research that improves understanding of human health, well-being and development.

Population Data BC has three layers of governance:
- The Governance and Oversight Committee which guides and supports the operations of Population Data BC. The membership consists of representatives of data stewards, research institutions and researchers;
- The Population Data BC Board which provides direction for achieving its mission as an organisation by directing and guiding the Executive Director in the development of Population Data BC's infrastructure development as well as its operations;
- The Operations Committee which plans and reviews the operational functions of Population Data BC, and ensures they are well coordinated amongst the partner universities.
Storage and Maintenance of Data Collections
Population Data BC holds a wide range of data collections at the University of British Columbia. Examples of the types of data in held are:

- Health care and health services data
- Population and demographic data
- Occupational data
- Early childhood data

The data collections are updated throughout the year and Population Data BC are continually working to expand the collections. A complete list of the data collections included can be found at http://www.popdata.bc.ca/data.

Data Linkage
The University of British Columbia performs both the linkage and storage and maintenance of data functions for Population Data BC. Fully identified data is received by Population Data BC and is separated into identifiable data such as name, address and date of birth; and content data such as diagnosis and interventions. The identifiable data is stored on a separate server to the content data. One or two named programmers use the identifiable data to update the Population Directory used to perform the linkage. In no cases are content data and identifiable data brought back together.

Facilitation of access to linked data
Population Data BC has developed a range of tools and services to assist researchers to access their linked data resources. These include:

- Website
  The Population Data BC website is comprehensive and includes information for researchers about data, application and approval processes, privacy and training and education;

- Metadata
  Population Data BC aim to provide as much information about the data collections and the data as possible on the website. They continue to systematically develop and expand their data documentation. They are currently developing an online documentation delivery system called MetaData Central;

- Coordination of applications and approvals
  An automated application tracking system has just been launched. “AppTracker” will enable staff to track projects from application date through the approval process and onto the completion of the project and link with or generate the required documentation e.g. approval letters;

- Assistance with project design;

- The Secure Remote Environment (SRE)
  The SRE is a central server accessible only via an encrypted Virtual Private Network through a firewall and use of two-factor authentication. It provides researchers with a central, secure place to store and analyse their data. A range of software is available in the SRE including SPSS, SAS and STATA. It
meets all the security requirements of the Data Stewards. Researcher can access the SRE from anywhere within Canada including at home;

- Training
The University of Victoria is leading the training component of Population Data BC. The main focus of the training program is the online Professional Specialization certificate in Population Health Data Analysis. There are also a number of other training activities including a program of workshops and events including an annual conference and a series of best practice and how to documents. Population Data BC also provides online privacy training, which is required before access is granted to the data. All researchers must complete an online privacy training course before they are given approval to access the data. If they fail to pass the test after two attempts they must attend a face to face training session.

Analysis and evaluation of linked data
Population Data BC does not conduct research.

Information Governance

- Legal and Regulatory Compliance
Population Data BC operates in a complex legislative environment which includes both Provincial and Federal legislation and regulations. In particular, British Columbia’s Freedom of Information and Protection of Privacy Act, E-Health Act and Personal Information Protection Act as well as the federal Privacy Act and the Personal Information Protection and Electronic Documents Act;

- Risk Assessment and Risk Management
Population Data BC works within a privacy risk management framework which describes a comprehensive approach to privacy which includes confidentiality agreements, privacy training, a Privacy Impact Assessment, a public website, accountability and advisory input, physical security, network security, and human resources controls, including the presence of a Privacy Officer;

- Audit, Reviews and Reporting Guidelines
Independent audits are conducted approximately every two years;

- Incident and Breach Management
Requirements for the management of incidents, breaches and complaints are outlined in policy and procedure documents as well as in data transfer and confidentiality agreements;
Conclusions and Recommendations

During my Fellowship I visited seven different data linkage organisations in three countries. All of these organisations, like the Population Health Research Network, are engaged in facilitating access to linked data from existing data collections to improve the understanding of human health, well-being and development. Each has different approaches to achieving this goal.

In order to achieve the potential benefits that access to linked data from across Australia can offer, the Population Health Research Network needs to build a data linkage network that:

- Minimises the barriers to accessing linked data;
- Minimises the risks to privacy and confidentiality; and
- Encourages and enables the use of data for the benefit of the Australian community.

During my visits I identified a range of solutions and approaches to minimising barriers to accessing linked data and minimising risks to privacy and confidentiality.

Approaches to minimising barriers to accessing linked data and encourage the use of data included the provision of:

- a wide range of data collections available for linkage which cover long periods of time and different subject areas e.g. health, education, justice;
- comprehensive metadata and data documentation about the data collections;
- simple, easy to understand application and approval processes;
- clear information about application and approval processes;
- staff to assist with application and approval processes;
- timely application and approval processes;
- timely data extraction processes;
- remote access laboratories (i.e. not having to go to a specific physical location for access);
- a range of software for analysis of data;
- training in data analysis techniques and data analysis software;
- dedicated staff to extract data for research;
- expert analysts to assist with research design and analysis.

Approaches to minimising risks to privacy and confidentiality included:

- the linkage function being undertaken by a separate organisation to the data storage and maintenance, facilitation of access and data analysis functions;
- the linkage function being undertaken by a separate group within the same organisation to the data storage and maintenance, facilitation of access and data analysis functions;
- a well-defined information governance framework;
- regular independent information governance and privacy audits;
• restriction of access to linked data for research to the staff of the linkage organisation;
• restriction of access to linked data to a physical location (e.g. at the linkage organisation or a satellite location);
• restriction of access to linked data to a virtual environment such as remote access laboratory;
• use of disclosure risk assessment tools;
• training for researchers in privacy and ethics;
• appointment of a dedicated privacy officer.

There are two parts of the data linkage process that can be the greatest barriers to timely access to linked data for researchers. These are the application and approval process and the extraction of the research datasets. In Australia the distributed data linkage model requires that each individual data custodian must extract the data for each research project. The time for several different data custodians to extract their data for a specific research project can be considerable. This is a different model to any of the organisations visited during my Fellowship. The units I visited all had some type of central repository with dedicated staff to extract data for research projects.

**Recommendation 1:** That the Population Health Research Network considers ways to minimise the amount of time and paperwork involved in the application and approval processes for access to linked data.

**Recommendation 2:** that the Population Health Research Network considers ways to streamline the data extraction process. Options to consider include data repository-type solutions as well as funding for dedicated staff for the data custodians.

All of the organisations visited during the Fellowship considered the protection of privacy and confidentiality a high priority. Some privacy strategies were common to almost all the units visited. These included:

• the separation of the linkage function from other functions either in a different organisation or within an organisation;
• the provision of training for researchers in privacy and ethics; and
• the appointment of a dedicated privacy officer.

**Recommendation 3:** That the Population Health Research Network mandate privacy and ethics training for all researchers accessing linked data through the Population Health Research Network infrastructure

**Recommendation 4:** That all Population Health Research Network participants consider establishing the role of information governance or privacy officer if the position does not already exist.
## Glossary

<table>
<thead>
<tr>
<th>TERM</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidentiality</td>
<td>The treatment of information that an individual has disclosed in a relationship of trust and with the expectation that it will not be used or divulged to others in ways that are inconsistent with the understanding of the original disclosure, without permission.</td>
</tr>
<tr>
<td>Consumers</td>
<td>Patients and potential patients, carers, organisations representing consumers' interests, members of the public who are targets of health promotion programs and groups asking for research because they believe that they have been exposed to potentially harmful circumstances, products or services.</td>
</tr>
<tr>
<td>Content data</td>
<td>The information in a record that contains person-specific information but does not contain identifiable data. Examples include hospital discharge codes, cancer diagnosis, cause of death, education results.</td>
</tr>
<tr>
<td>Data custodian</td>
<td>The organisation or agency which is responsible for the collection, use and disclosure of information in that dataset. The data custodian is responsible for contributing to the guidelines and approval processes on the use of the data, including involvement with ethics committees and input to the protocols surrounding data use.</td>
</tr>
<tr>
<td>Data dictionary</td>
<td>A tool that describes the meaning, relationships to other data, origin, instructions for usage and format of data elements within a data collection.</td>
</tr>
<tr>
<td>Dataset</td>
<td>A set of data that has been collected through a one-off research process, or a subset of data that has been assembled by extracting those records or parts of records from one or more data collection(s) that the researcher has permission to see.</td>
</tr>
<tr>
<td>ACRONYM</td>
<td>DEFINITION</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>ALF</td>
<td>Anonymous Linking Field</td>
</tr>
<tr>
<td>HIRU</td>
<td>Health Informatics Research Unit</td>
</tr>
<tr>
<td>HTTPS</td>
<td>Hypertext Transfer Protocol Secure</td>
</tr>
<tr>
<td>ICES</td>
<td>Institute for Clinical Evaluative Sciences</td>
</tr>
<tr>
<td>ISD</td>
<td>Information Services Division NHS Scotland</td>
</tr>
<tr>
<td>MCHP</td>
<td>Manitoba Centre for Health Policy</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>OAHP</td>
<td>Ontario Agency for Health Protection and Promotion</td>
</tr>
<tr>
<td>OHIP</td>
<td>Ontario Health Insurance Plan</td>
</tr>
<tr>
<td>PHIN</td>
<td>Personal Health Information Number</td>
</tr>
<tr>
<td>SAIL</td>
<td>Secure Anonymised Information Linkage</td>
</tr>
<tr>
<td>SHIP</td>
<td>Scottish Health Informatics Programme</td>
</tr>
<tr>
<td>SHIS</td>
<td>Scottish Health Information Service</td>
</tr>
<tr>
<td>SHIS-R</td>
<td>Scottish Health Information Service - Research</td>
</tr>
</tbody>
</table>