



Office of the Chief Statistician and Performance
4N-06 St Andrews House
Regent Road
Edinburgh
EH1 3DG
Scotland
UK

A Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles

Introduction

The Centre for Data Linkage (CDL) welcomes the opportunity to make a submission in response to the consultation paper on '*A Scotland-wide Data Linkage Framework for Statistics and Research: Aims and Guiding Principles*'.

The Centre for Data Linkage (CDL) is the national component of the Population Health Research Network (PHRN) and was established in April 2009 as a unit within the Faculty of Health Sciences at Curtin University. Hosted by the Curtin Health Innovation Research Institute (CHIRI), the CDL has established a secure data linkage facility, and to enable linkage between jurisdictional datasets, and between these datasets and research datasets using demographic data.

Relevant CDL capability

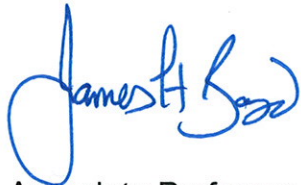
A team of data linkage experts has been established at CDL and are actively involved in the design, development and testing of this state-of-the-art national data linkage facility. In addition to its important role as a secure national data linkage facility, the CDL is also responsible for:

- research and review of various linkage models and technology landscapes;
- assisting with the secure delivery of national linked data to researchers;
- providing technical advice and assistance to other data linkage units and key committees;
- review and development of data quality assurance tools.

Consultation questions

A detailed response to the consultation paper is attached in the respondent information form and draws on the CDL's expertise in the data linkage area. We hope this provides some useful information from an international perspective and would be pleased to discuss any of the matters raised in our submission in further detail, at your convenience.

Yours sincerely



Associate Professor James Boyd
Director, Centre for Data Linkage
Curtin University

12 June 2012

CONSULTATION QUESTIONS

**Are you responding *primarily* as a data custodian, data user or data subject?
(We recognise all people are data subjects and many organisations act as data guardians and data users, but please tick only one box)**

- Data Custodian
- Data User (e.g. researcher)
- Data Subject (e.g. member of the public or group representing citizens)

Centre for Data Linkage

The Centre for Data Linkage (CDL) is the national component of the Population Health Research Network (PHRN) and was established in April 2009 as a unit within the Faculty of Health Sciences at Curtin University. Hosted by the Curtin Health Innovation Research Institute (CHIRI), the CDL has established a secure data linkage facility, and to enable linkage between jurisdictional datasets, and between these datasets and research datasets using demographic data.

Expertise in data linkage

A team of data linkage experts has been established at CDL and are actively involved in the design, development and testing of this state-of-the-art national data linkage facility. In addition to its important role as a secure national data linkage facility, the CDL is also responsible for:

- research and review of various linkage models and technology landscapes;
- assisting with the secure delivery of national linked data to researchers;
- providing technical advice and assistance to other data linkage units and key committees;
- review and development of data quality assurance tools.

For more information on Data Linkage at CDL please contact:

Anna Ferrante (a.ferrante@curtin.edu.au) or
James Boyd (j.boyd@curtin.edu.au).

1. Are there any benefits of data linkage for statistical and research purposes that are not sufficiently described here?

Yes, there are further benefits No, the benefits are described fully

If you ticked 'yes', please describe the further benefits of data linkage for statistical and research purposes.

There is substantial public benefit to be gained from research using linked data. As discussed in the consultation paper, secure data linkage technologies offer opportunities to maximise the value of administrative data and will provide an important resource for both government and university researchers.

Data linkage can provide valuable information that would not be available any other way. Using data linkage techniques, it is possible to analyse relationships among variables without collecting additional data. Linkage therefore increases the value of data that are already collected and reduces the need for additional data collection.

Once the linkage infrastructure is in place, it will enable the creation of complex linked data and allow new and innovative research that can be used to enhance the delivery of public services. The benefits of such a programme would significantly improve research capabilities in Scotland and provide a robust platform to inform policy.

Some specific examples of where data linkage has been used to great effect across a range of health and health related areas can be found on the PHRN website (www.phrn.org.au/about-us/what-we-have-learnt).

Research studies using similar national and international infrastructure have contributed significantly to both services and outcomes. These include:

Data Linkage Australia (DLA) – Australia

The Data Linkage Australia (DLA) draws on the unique competitive advantages possessed by Western Australia in its ability to attract external science investment, due to its possession of comprehensive, population-based health databases, and recent local achievements in community leadership and medical informatics that have created individual identity-relationships between millions of database records. The collaboration brings together a multi-disciplinary group of science and technology leaders who have developed infrastructure that supports population based research (and research into cross portfolio development pathways).

Population Health Research Network (PHRN) – Australia

The Population Health Research Network (PHRN) has been established to develop data linkage units in all States/Territories around Australia supporting

nationally and internationally significant population-based research that will improve health and enhance the delivery of health care services in Australia.

The PHRN comprises a Program Office located in Perth, Western Australia, a Centre for Data Linkage (CDL) at Curtin University in Western Australia, a Remote Access Laboratory at the Sax Institute in New South Wales and a network of Project Participants and Data Linkage Units located in each Australian state/territory (including the Data Linkage Branch in the Western Australian Department of Health and SANT DataLink).

The PHRN has the objective of building a national infrastructure for the development and promotion of data linkage for population and clinical health datasets for research purposes. This network includes national elements including the National Linkage System (Centre for Data Linkage, Curtin University) and the Secure Unified Research Environment (Sax Institute, Sydney).

Centre for Data Linkage (Curtin University)

The Centre for Data Linkage (CDL) is the national component of the PHRN and was established in April 2009. Hosted by the Curtin University, the CDL has established a secure data linkage facility, and to enable linkage between jurisdictional datasets, and between these datasets and research datasets using demographic data.

The CDL does not hold full datasets; it links the demographic data that has been separated from the remainder of each dataset to create 'linkage keys' (clinical or service information are not needed by the Centre for Data Linkage and is not provided to it).

In addition to its important role as a secure national data linkage facility, the CDL is also responsible for:

- research and review of various linkage models and technology landscapes;
- assisting with the secure delivery of national linked data to researchers;
- providing technical advice and assistance to other data linkage units and key committees;
- review and development of data quality assurance tools.

Australian Institute of Health and Welfare (AIHW)

In the context of the new Commonwealth data linkage framework, AIHW has expanded its capabilities in data linkage by becoming the first organisation accredited to undertake complex data linkage projects involving Commonwealth data. This accreditation gives assurance to data custodians and the public that high levels of security and strong governance processes are in place to protect the privacy of individuals, while maximizing the potential of available datasets

held by the Institute to provide new insights into complex policy issues, thereby supporting better policy decisions.

SAIL Databank – Wales

The Secure Anonymised Information Linkage (SAIL) Databank brings together the widest possible range of person-based data using robust anonymisation techniques for health related research. It is operated by the Health Information Research Unit (HIRU) based at Swansea University. HIRU aims to realise the potential of electronically-held, person-based, routinely-collected data to conduct and support health-related studies. The Databank holds over 500 million records and continual growth is in progress.

SAIL has been designed and developed to: 1) ensure data transportation is secure; 2) operate a reliable record matching technique to enable accurate record linkage across datasets; 3) anonymise and encrypt the data to prevent re-identification of individuals; 4) apply measures to address disclosure risk in data views created for researchers; 5) ensure data access is controlled and authorised; 6) establish methods for scrutinising proposals for data utilisation and approving output; and 7) gain external verification of compliance with Information Governance.

2. Are there challenges or barriers preventing more effective and efficient data linkages for statistical and research purposes taking place that are not sufficiently described here?

Yes, there are further challenges No, the challenges have been identified

If you ticked 'yes', please describe the challenges or barriers.

The methods and techniques around data linkage are well established and the new developments (exploiting advances in technology) have the potential to improve timeliness and efficiency. The Data Linkage Framework for Scotland should build on existing national and international collaborations, infrastructure and skills – it should leverage these developments to 'fast track' the programme.

The main challenges or barriers to be resolved in realising the potential of the Data Linkage Framework for Scotland include:

- Legislation – many datasets that could make a major contribution will be subject to specific legislation that defines the conditions of data release and/or use. The extent of this type of legislation and its complexity can create difficulties of interpretation with regard to the release of data for linkage projects;

- Security, ethics and privacy – in addition to legal requirements, access to many datasets may also be subject to privacy and ethical review. The design of any developments need to enable safe and secure access to linked data (i.e. that they apply the appropriate governance process, protocols and standards) to address any privacy and confidentiality concerns and facilitating the use of this data within the constraints of privacy, confidentiality and the law;
- Capacity and access – the operations required to fulfil a data request can pose a substantial workload. The infrastructure involved in undertaking data linkages need to be scalable, fast and efficient to ensure timeous response to important policy and research questions (this includes application, specification, review, workflow, linkage, metadata and analysis);
- Expertise – data linkage requires expertise in three broad areas – knowledge of the datasets available for linkage and their characteristics and limitations, skills in data linkage methods and skills in analysing linked information. By itself, a basic level ability to use available linkage software is insufficient, because correct interpretation of a linked datasets depends on an understanding of the structure and content of and variation within the component datasets;
- Researcher engagement - it is very often time consuming to establish data linkage projects in terms of legal and governance arrangements. A process that manages/streamlines all the processes involved in executing a linkage project (from start to end) would ensure effective and efficient data linkages (these processes need to address any uncertainty around legal, privacy and data custodian constraints);
- Cost effective national infrastructure – development and maintenance of the data linkage infrastructure should be relatively inexpensive (and can leverage national and international experience).

3. Are the guiding principles sufficient and appropriate? Please explain your answer fully and make suggestions for improvement.

Yes, they are sufficient and appropriate No, they are not

Please explain your answer fully and make suggestions for improvement.

The broad objective of the 'Data Linkage Framework' is to facilitate data linkage as an efficient mechanism for obtaining information from existing data collections that can inform policy and improve public services. However, the design of the infrastructure should also include mechanisms to achieve adequate control and manage risk including the use of specified standards and audits against those standards, and the potential application of sanctions should be considered.

The framework could consider the following areas to be included in the principles:

- Developing a strong Information Governance framework throughout the infrastructure to ensure the required safeguards for, and appropriate use of, personal information;
- Developing minimum standards for the secure management of information and Information Governance including ISO/IEC Information Security standards (covering: Information technology/Security techniques/Information security management systems);
- Management of all aspects of the data linkage process from data access to analysis to disposal (and risk of privacy and/or confidentiality breaches) for research projects ensuring/considering:
 - Confidentiality and privacy
 - Data Protection
 - Information security
 - Record Management (Data retention/destruction policies are applied)
 - Data access agreements
 - Data analysis
- Developing procedures and guidelines regarding the usage and archiving of linked datasets, access to linked datasets and intellectual property created through the analysis of linked datasets.

4a. Are the objectives set out for a Privacy Advisory Service in Section 3c the right ones?

Yes, the objectives are right

No, they are not

Please explain your answer fully and make suggestions for improvement.

The Privacy Advisory Service should provide appropriate expertise and resources, to resolve the governance issues that currently hamper worthwhile data linkage projects which are likely to benefit individuals and/or the community. It needs to support data linkage projects, providing advice and guidance for custodians and researchers on the data linkage process, and enhancement of researchers' awareness of the legislative, policy and ethical aspects of data linkage.

The creation of a Privacy Advisory Service addresses some of the challenges associated with facilitating data linkage research and provides a focal point for ethical, legal and social issues associated with data linkage projects.

4b. Do you wish to be consulted on firmer proposals for a Privacy Advisory service as and when they are developed?

Yes No

5a. Are the functions that will be led by the National Data Linkage Centre set out in section 3d the right ones?

Yes, they are the right functions No, they are not

Please explain your answer fully and make suggestions for improvement.

The National Data Linkage Centre in Scotland needs to provide a high-performance data linkage environment with enhanced speed and storage capacity. The Centres functions/capabilities should also include:

- Information and communication technology support;
- The acquisition and maintenance of data linkage research equipment;
- Data linkage expertise and the development of systems, technologies and methods;
- Standardisation of components of datasets, such as metadata that facilitate data linkage;
- Workforce training (data linkage and analytical) with training programs for workforce development;
- Data management and analytical capacity (developing skill in statistical analysis of complex linked datasets); and
- Mechanisms for governing and managing data linkage capacity.

5b. Do you wish to be consulted on firmer proposals for a National Data Linkage Centre as and when they are developed?

Yes No