

POPULATION HEALTH AND CLINICAL DATA LINKAGE

Introduction

This document will form the basis for the development of an investment plan for the Population health and clinical data linkage capability by a facilitator. While the document provides an outline for a national capability, a number of issues have been raised during the scoping of the capability that remain to be considered and discussed in-depth. In addressing the issues, the facilitator will need to recognise the health research leadership role of the NHMRC and the national health information leadership role of the AIHW.

Description

Data linkage is the bringing together from two or more different sources data that relate to the same individual, family, place or event. Australia is well placed to be a world leader in health research using data linkage because of its rich collection of large population databases, its favourable policy and legislative environment, and its existing foci of research excellence.

Building a national capability for research using data linkage will require investment in structures and frameworks for governance, privacy, ethics and community involvement; information technology and information management; data sets and metadata; methods and tools for data linkage and analysis of linked datasets; and human capacity.

Research applications of data linkage include studies of:

- Use and costs of health services, for example studies of the patterns and costs of hospital care and out-of-hospital care for people with specific conditions.
- Outcomes of different models of health services delivery, through linking data about service use with death and disease registers.
- Outcomes of specific clinical and therapeutic interventions, for example survival post-surgery, or post-marketing surveillance for adverse effects of drugs.
- Disease causation, for example questionnaire information from subjects in cohort studies can be linked with disease registers to allow incident health conditions to be related to baseline information on risk factors and pre-existing conditions.
- Relationships among social factors and health throughout the life course, through integration of health data with data from other sectors, including aged care and community care services, justice and education.
- Relationships among environmental factors and health, through linking individual-level health data with data on environmental exposures for specific geographic units.
- Associations between biological factors and health outcomes, through linking disease registers with genetic and molecular data sets.
- Long-term follow-up of outcomes for subjects in clinical trials, through linkage with death and disease registers.

Rationale

The 2004 Investment of Review of Health and Medical Research¹ (Grant Review) identified that investment in policy and practice focused research can deliver significant returns to the Australian health system, and improve the effectiveness of the \$66.6bn annual spend on health care. Investment in national capability for data linkage will make an important contribution to

¹Australian Government. *Sustaining the virtuous cycle for a healthy, competitive Australia. Investment Review of Health and Medical Research*. Final report. Canberra: Commonwealth of Australia, 2004

addressing this area as data linkage has broad applications in public health and epidemiological research, health services research and clinical research. Linkage of population-based data sets facilitates research into policy-relevant issues spanning prevention, detection and management of disease, use and costs of health and community services, and the long- term health and societal outcomes that these achieve.

The potential benefits of data linkage include:²

- Increased cost-efficiency of research compared with performing de novo longitudinal studies and other more traditional approaches to epidemiologic and health services research.
- Adding value to existing information assets and generating a research return on the substantial existing investment in routine administrative and clinical data sets within health.
- Allowing integration of data from the health sector with a wide range of data from other sectors (including, for example, education, community services, police, justice, planning, transport), without the requirement for unique personal identifiers across these sectors.
- Conserving the privacy of individuals by reducing the need for release of names and other personal identifiers to researchers.
- Contributions to medical and scientific knowledge in the form of scientific publications, conference presentations, and theses and dissertations prepared by research trainees.
- Fostering collaborative research involving population health researchers, clinical researchers and biomedical scientists, with direct benefits for clinical outcomes.
- Community development through enhancing interactions among researchers, clinicians, administrators, consumer groups and the mass media.

A world-leading data linkage system (the Western Australian Data Linkage System) has been operating since 1995. Over the period 1995 to 2003, this system supplied data for 258 projects, which produced 708 research outputs, including 172 journal articles. Research using linked data from the Western Australian Data Linkage System was directly responsible for a range of changes to policy and clinical practice.³

Another state based system using a similar approach has been established to service New South Wales and the Australian Capital Territory (the Centre for Health Record Linkage).

National and state agencies and researchers are undertaking a variety of other data linkage activities, for example, both the Commonwealth and Victorian State Government have invested in a federated data grid which links clinical and genomic data (Bio21:MMIM⁴). The Commonwealth Scientific and Industrial Research Organisation (CSIRO) has been developing service-oriented software for linking data electronically in a secure environment. The New South Wales Government and non-government partners have invested in the 45 and Up Study⁵, the largest follow-up study of health ever conducted in Australia, which will recruit 250,000 people and track their health using data linkage.

Australia is therefore well-placed to consolidate its position as an international leader in research using data linkage.

² Data Linkage Australia. *Scoping paper: a model for a data linkage facility in New South Wales*. Sydney: the Sax Institute, 2005

³ Brook EL, Rosman DL, Holman CDJ, Trutwein B. *Summary report: Research outputs project, WA Data Linkage Unit (1995-2003)*. Perth: WA Data Linkage Unit, 2005.

⁴ <http://mmim.ssg.org.au/>.

⁵ <http://www.45andup.org.au/>

The current Australian context offers several unique opportunities for building a national capability:

- The large number of high-quality health data sets, describing most aspects of the Australian health system in substantial detail.
- Australia's privacy laws and the privacy principles on which they are based provide a firm foundation for data linkage initiatives, and protection of individuals' privacy.
- The Australian population is relatively large, and very diverse, compared with the populations covered by other well-developed data linkage systems (such as Manitoba, British Columbia and Oxford). Scotland is currently the only country that can perform a wide range of population health studies using linkage at the national level.
- Practices and policies for data linkage are already tested and in place in several Australia jurisdictions.
- National health data sets are endorsed by all nine Australian Governments through the Australian Health Ministers' Advisory Council's (AHMAC) National Health Information Group.
- Consistent metadata specifying these national collections is available in the National Health Data Dictionary and the Australian Institute of Health and Welfare's (AIHW) electronic repository METeOR.⁶
- The Statistical Information Management Committee (SIMC), a subcommittee of the National Health Information Management Principal Committee, has recently completed a project to document the state of readiness of jurisdictions to create linkages between health data sets for research purposes, and is currently working on a framework for linkages at the national level.
- The Australian Bureau of Statistics (ABS) will allow limited linkage to census data, and thus a wide range of new social and economic variables, from the 2006 Population Census onwards.
- The Australian Research Council has provided funding for important initiatives, such as the ARACY/ARC Research Network: Future Generation.

The opportunity now exists to realise Australia's potential competitive advantage in research using linked population and clinical data sets. Investment in, and rapid development of, a national capacity could see Australia positioned as a 'population laboratory' for the world, and attract international investment in Australian research, particularly in health services research, cohort studies, and pharmaceutical research including post-marketing drug surveillance.

Principles

There is broad support from stakeholders for the development of a national data linkage capability and agreement that the principles underpinning such a capability include:

- Linked data are to be used for research and statistical purposes only, not for the clinical care of individuals, or regulatory or enforcement purposes.
- The identity of individuals should never be disclosed.
- There should be arrangements for consumer knowledge of, and involvement in, data linkage arrangements.

The capability will need to be developed and operated in a manner consistent with the NCRIS principles which encourage research excellence, collaboration and accessibility. The principles are set out in the NCRIS *Roadmap*. In general, the principles require the research infrastructure in the capability to be widely accessible for use by researchers on the basis of

⁶ <http://meteor.aihw.gov.au>

merit. To achieve this, the capability will need suitable governance and management arrangements that will encourage collaboration within the research community and with research users.

Infrastructure/Support Required

The primary aim of national infrastructure should be to make linkable data across jurisdictions **available and useful to researchers**. This includes not only mechanisms for linking records across jurisdictions, but also infrastructure for making these data accessible for research. This poses particular challenges because the data sets involved are potentially massive in size, highly heterogeneous, use different and sometimes unspecified terminologies, and are distributed among different owners and governed by different legislations.

As mentioned above, Australia already has highly-developed, but localised, systems for linkage of population and clinical data sets. However, substantial investment in infrastructure is needed to move to full national population coverage, and to facilitate linkage with national-level population data sets. Areas requiring investment include structures and frameworks for governance, privacy, ethics and community involvement; information technology and information management; data sets and metadata; methods and tools for data linkage and analysis of linked datasets; and human capacity.

Specific investment in infrastructure for research will need to be supported by a range of actions from the **health sector**, with leadership from the National Health and Medical Research Council (NHMRC). Key among these will be:

- High level commitment, from Australian Health Ministers, to the value of linked health data.
- Cooperation from data custodians to provide data for scientifically sound and ethically approved research.
- Simplification of ethics approvals processes to minimise the need for multiple ethics approvals.

In terms of investment in **research infrastructure**, development should be firmly based on what has already been put in place, or is underway. Specific requirements include:

1. A national **health data linkage infrastructure** comprising State-based units and a national unit:
 - The national unit will develop jurisdictional datasets, such as deaths and hospital separations, into national linkable datasets.
 - The national unit will facilitate approval of data custodians for access to the linkable datasets, and ensure that release of Commonwealth data is for research projects demonstrated to be in the public interest.
 - The national unit will perform approved linkage with national datasets, where these are available in a linkable form (including Medicare Australia data, deaths, hospital morbidity and cancer registrations).
 - The State-based units will provide linkage services relating to state-specific projects, unique jurisdictional data holdings, and local research data sets.
 - Both State-based and national units will potentially have roles in facilitating the linkage of a range of existing research data sets and registers that are held by hospitals and non-government custodians.
 - As well as undertaking data linkage, both State-based and national units will provide infrastructure to support access to, and use of linked data by researchers (including tools and services for authorisation, authentication and access, certification, query and search, and data analysis and quality control).

2. **A national centre for health data linkage development**, covering:
 - Coordination of the national health data linkage system.
 - Policy development, including work with NHMRC to ensure adequate recognition of research based on health data linkage and development of robust and efficient ethics approval processes.
 - Negotiation for access to data sets and for database structures that assist data linkage.
 - Methodology standardisation and development, potentially including national software development/acquisition.
 - Development and maintenance of a data set catalogue and data set documentation.
 - Development and maintenance of metadata descriptors, repositories and business glossaries (leveraging emerging developments in ontologies and data standards).
 - Gathering and sharing intelligence relating to emerging methods and technological solutions, including service-oriented architectures, privacy-preserving analytics and methods to handle non-structured data.

3. **Human capacity building**, including:
 - A funded national plan to deliver the people required to support and develop the national research capacity.
 - Training for existing researchers in methods for data linkage and analysis of linked data sets.

Issues to be resolved through the facilitation process

The issues include:

- In addition to leadership from the NHMRC, what, if anything, is required to facilitate health sector engagement with, and support for, development of the capability?
- To what extent is the model of State-based units, a national unit, and a centre for health data linkage development supported? Assuming general acceptance of the model:
 - What national governance and management model is appropriate?
 - Should the national health data linkage unit and the national centre for health data linkage development be co-located? Should either or both these roles be allocated to a new body or to an existing body or bodies?
 - What are the most appropriate roles for the NHMRC and the AIHW?
 - Should a national health data linkage unit be based on the Western Australia Data Linkage infrastructure?
 - How will new State-based units be identified, and the relevant stakeholders engaged to ensure their establishment?
 - How will the model interface with emerging developments including Bio21:MMIM and the work of CSIRO's e-Health research centre?
- To what extent is the model of separation of linkage key development and maintenance, from the research data supported? (Legislative protections can provide strong alternative arrangements as in the AIHW and ABS) .
- Should consideration be given to a data warehouse/repository, a model whereby health data remains with distributed custodians, or a federating model of data – a virtual repository? Or could a hybrid approach work, with some linkage keys available through a state or national centre and some locally, and some data available through a central repository and some virtually?
- What are the current opportunities relating to emerging technologies?
- How should this capability align with the work of the National e-Health Transition Authority (NeHTA), including its standard and guidelines on user authentication and

access control measures? What will be the effect of the expected introduction of NeHTA's Individual Healthcare Identifier (IHI) in late 2007?

- How will consumer perspectives be acknowledged and addressed in the development and operation of the capability?
- What new human capacity will be required to support the capability?
- What, if any, initiatives can be expected to impact on the capability after the development of the investment plan?