

Submission

2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Dr Merran Smith
Title/role	Chief Executive
Organisation	Population Health Research Network

Introduction

The Australian Government has requested submissions on the National Research Infrastructure Capabilities Issues Paper. Responses to the Issues Paper will inform development of the 2016 National Research Infrastructure Roadmap (Roadmap 2016).

This submission is from the Population Health Research Network (PHRN). It includes responses to specific questions (Questions 1-17, 24-27, and 30-35 inclusive) in the Issues Paper. It also includes a discussion on future directions and research infrastructure capability requirements for health and medical sciences in Australia over the next decade, with particular reference to large population-based, unit record (person level) data and related priority data.

Question 1: Are there other capability areas that should be considered?

The capability areas identified in the Issues Paper cover Australia's science and research priority areas. It is important that the Health and Medical Science capability area includes the broader social determinants of health. Science and research related to other human services including education, justice and families are relevant to Health and Medical Science. Alternatively they could be considered under Understanding Cultures and Communities.

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

There are opportunities for coordination within and between capability areas. PHRN has had productive discussions with other groups within the Health and Medical Science areas. In terms of an optimum governance model, most capabilities do exhibit the characteristics identified in Section 3.2 of the Issues Paper. In terms of greater coordination within broad capability areas, an overarching Council made up of senior governance/ management from within each area could be a vehicle to achieve greater coordination, noting that some current NCRIS capabilities cover more than one broad capability area.

Questions 3 and 4: Should national research infrastructure investment assist with access to international facilities? What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

At this stage, there is no demand or need to send Australian health and human services data to international facilities for linkage. However, requests from Australian researchers to access linked data from other nations and for international researchers to access linked Australian data are likely to increase over the next decade. These international collaborations should be encouraged. The need for specific infrastructure funding to support these collaborations will depend on the nature of the collaboration. PHRN is a member of the International Population Data Linkage Network so is well placed to know of developments.

Question 5, 6 and 7: Should research workforce skills be considered a research infrastructure issue? How can national research infrastructure assist in training and skills development? What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

Research workforce skills are a research infrastructure issue both for the operation of research infrastructure and in the development of researchers with the skills to use the infrastructure. For example, there is a need for more researchers/analysts skilled in using health and medical sciences data. Training, education and skills development are primarily the responsibility of universities, and research institutes through undergraduate and post-graduate training, other training programs as well as on the job training. Research workforce skills could be considered a critical issue in harvesting additional benefit from use of the research infrastructure. National research infrastructure capabilities could play a role in skills development through specific training programs. However, the limited funds available for national research infrastructure going forward mean that other support is likely to be required. Strategies should be developed by governments in consultation with educational institutions and others to address training requirements. While this section focuses on the research workforce, it is important to remember that other approaches will be required to ensure the skilled workforce needed to operate Australia's national research infrastructure.

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

Publicly funded research infrastructure should be made widely available to publicly funded researchers with as few barriers as possible to access. Priority of access should be based on merit including as appropriate:

- Scientific merit ;
- Public interest;
- Compliance with the relevant legislation and regulations.

Private organisations should not be excluded from access but should meet the access criteria and pay reasonable costs for access.

The PHRN supports merit-based researcher access to Australia's national data linkage infrastructure. Because much of the data used by the infrastructure is drawn from statutory and administrative data which is collected across the population without consent, researcher use must have a public benefit. PHRN facilities generally have or are moving towards a partial cost recovery model. A number of research projects supported by the PHRN infrastructure have international collaborators. These collaborations are valued and expected to continue. Generally linked data is not directly

available to private sector organisations. However, such organisations can commission research to be undertaken by approved research groups under agreed conditions. Charges will depend on the nature of the project.

Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?

Defunding/decommissioning may impact a facility, an activity or a whole capability. The value of the infrastructure, the assets it has generated, the skills, knowledge and experience of those who operate it, and alternative arrangements to address the needs of its users should be considered in any decision-making process related to defunding. The possibility of transitioning to alternative arrangements should also be considered. As the Issues Paper points out, withdrawal of Australian government funding may result in withdrawal of co-funding. Future projected demand and the extent to which other countries are investing in similar infrastructure may provide further guidance.

Question 10: What financing models should the Government consider to support investment in national research infrastructure?

Australia's approach to funding national research infrastructure through the NCRIS programs has been successful. In PHRN's case, the collaborative approach has enabled significant cash and in-kind co-investment around 1:1.5 from state and territory governments, government agencies, universities and research institutes. It has also reduced waste by enabling implementation of agreed infrastructure without costly duplication.

The Australian Government announcement of ongoing operational funding for the existing NCRIS network is very welcome. The move away from a short term (annual) funding horizon to a longer term focus will support a more strategic investment cycle and provide for some operational efficiencies. However capital funding has not been provided. Funding constraints associated with the current global economic cycle are well understood and these limit the capacity of government and other sectors of the economy. The scale of capital expenditure required to support capabilities varies as do the funding options. In PHRN's case, the public benefit which underpins the data linkage infrastructure may limit commercial funding opportunities. In several other developed countries governments are directly providing significant capital (and operational) support for data linkage infrastructure (see Health and Medical Science section for further details). One option that has been suggested in the Australian health research infrastructure context is for some proportion of health and medical research funding to be allocated to capital expenditure.

Question 11: When should capabilities be expected to address standard and accreditation requirements?

It is considered important for national research infrastructure to meet agreed standards. Meeting national/international standards is particularly important if national/international comparisons are being made or results are being used for the development of commercial products. In PHRN's case, some Australian and/or international standards may be applied to aspects of the work. In other cases, standards need to be developed. The PHRN is already progressing development of national data linkage standards. In our experience, the development of standards requires time and

expertise, as well as a level of maturity of the infrastructure. External review through accreditation or some other appropriate process would be worthwhile, where it is cost effective.

Question 12, 13 and 14: Are there international or global models that represent best practice for national research infrastructure that could be considered? In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined? Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?

In terms of data linkage the FARR and Administrative Data Research Network models in the UK would be worth considering, see below for more details.

Health and Medical Science (Questions 15-17)

Research priorities 2017-2027

The health priority in the Australian Science and Research Priorities¹ is *“Build healthy and resilient communities throughout Australia by developing treatments, solutions and preventative strategies to improve physical and mental well-being and improve the efficiency and effectiveness of Australia’s health care system. This priority is focussed mainly on health services and public health research.”* To achieve this, research in three areas will be required:

- understanding health and disease;
- improving health;
- improving the health system.

Access to high quality and/or linked data will be an essential for research in all three priority areas.

Understanding Health and Disease

The World Health Organisation defines health as *“a state of complete and physical, mental and social well-being and not merely the absence of disease or infirmity”*². Understanding health and disease in this context means understanding not only the complex biological interactions of the human body but also the complexity of human social, emotional, lifestyle and environmental interactions both at the individual and population levels.

Many health and social problems are the result of these complex interactions (e.g. obesity; drug and alcohol abuse), and will require detailed knowledge of the causal pathways that result in these

¹ Australian Government Science. Health Capability Statement [Internet]. Canberra ACT: Australian Government Department of Industry Innovation and Science [updated 2015 November; cited 2016 Aug 14]. Available from:

<http://www.science.gov.au/scienceGov/ScienceAndResearchPriorities/Pages/Health.aspx>

² Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19 June - 22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

outcomes in order to find ways to both prevent and treat. A range of research methods and advanced analysis techniques will also be required to understand the complex interactions between the genome, proteome, metabolome, environment and phenome. These techniques will include:

- Phenome-wide association studies³ which investigate the influence of one or more genetic variants across a range of different human phenotypes.
- Genome-wide association studies.
- Population level epidemiology using a wide range of linked data collections, including data collections outside of health.

Improving Health

The development of new treatments and preventive strategies will be required to improve health at both the individual and population levels. Methods to more quickly translate research findings into clinical practice and improved health and well-being outcomes will also be important.

Precision medicine is a term used to describe the situation where a person's healthcare is tailored to the individual based on their genetics, environment and lifestyle. Advances in all aspects of human biology and medicine in recent decades, including the sequencing of the human genome and advances in cell biology, biochemistry and imaging technology, mean that precision medicine is now a possibility in the foreseeable future. There will be considerable research activity in this area in the next ten years.

Improving the Health System

Australia has a complex health system which is funded and run by the Commonwealth, state and territory governments and the private sector. Total expenditure on health was estimated at \$154.6 billion in 2013–14⁸. This was 9.8% of gross domestic product. The rising prevalence of chronic conditions multi-morbidity and obesity is challenging conventional approaches to delivery management and financing of health care. The social, economic, environmental and geographic determinants of these emerging threats to the quality of life and lifespan of Australians need to be understood to provide appropriate and cost-effective services.

Research in required to:

- Better understand health systems and the causes of variations in health care.
- Develop and evaluate innovative solutions to the equitable delivery of health care e.g. regional and remote; indigenous health; personalised medicine; rare diseases etc.
- Develop effective health promotion and protection strategies and best practice models of care.
- Determine the impact of medicines and therapies using real world data.

The value of access to personal information for research

National and International perspectives

Across the world governments have been acknowledging the absolute necessity of access to high quality administrative data and other big data for health, medical, social and policy relevant research. Some countries such as the United Kingdom have invested heavily in infrastructure for

data linkage and management. Examples include the FARR Institute³⁴ and the Administrative Data Research Network initiatives⁵.

The Productivity Commission devoted a whole chapter in its 2012-13 Annual Report⁶ to the use of linked administrative data to achieve better policy outcomes. It is currently in the middle of an inquiry in to Data Availability and Use. The Issues Paper includes specific questions about how to overcome barriers and improve access to linked data. The Senate Select Committee on Health recently released its report entitled “Big health data: Australia's big potential”⁷ which highlighted the potential of using linked data to inform health policy and improve health services. The Issues Paper also highlights the need and value of nationally linked research and administrative data.

There is no doubt that high quality, nationally linked data is essential to the future of health and medical research in Australia. However, there is often confusion about the difference between open data initiatives and linked data initiatives.

Open data is data that anyone can freely access, use, modify, and share for any purpose. Open data usually refers to aggregate data or non-sensitive data e.g. the My Hospitals website.

This level of data is not sufficient to support investigation of the complex interactions and biology necessary to understand the causes of disease, develop new treatments or to understand and improve our complex health system. Access to unit record data and sometimes personal data is required for these purposes.

Privacy and Confidentiality

A key challenge in making personal or unit record data available for research is the need to balance the right of people to control their personal information⁸ with the competing public interest in the research⁹. This juggling act in combination with Australia’s federated system of government has resulted in a complex and multi-layered legal framework which includes privacy protecting principles and provisions in three different bodies of law.

There is strong consumer and community interest in both the protection of privacy and the benefits of research. Therefore, access to population-based, unit record data must be managed carefully to ensure that it is only used for research for the public benefit and risks to individual privacy are minimised. There is considerable enthusiasm in the research community for the establishment of large, linked research data repositories of data collected without consent for research with minimal

³ <http://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/initiatives-in-informatics-research/> [accessed 29/07/2016]

⁴ <http://www.mrc.ac.uk/news/browse/20-million-for-new-health-informatics-research-institute/> [accessed 29/07/2016]

⁵ Economic and Social Research Council. Annual Report and Accounts 2013-14. <http://www.esrc.ac.uk/news-events-and-publications/publications/corporate-publications/annual-report-and-accounts/>[accessed 29/07/2016]

⁶ Productivity Commission 2013, Annual Report 2012-13, Annual Report Series, Productivity Commission, Canberra.

⁷ Senate Select Committee on Health. May 2016. Sixth Interim Report, Big health data: Australia’s Big Potential.

⁸ United Nations International Covenant on Civil and Political Rights Article 17.

⁹ United Nations International Covenant on Economic, Social and Cultural Rights Article 15.

approval requirements for access. However, a compromise between ease of access to and data available from such an environment would likely be necessary. Safeguards could include some aggregation of data item categories and agreement on the types of research that could be conducted using the data.

Linked Data in Australia 2017-2027

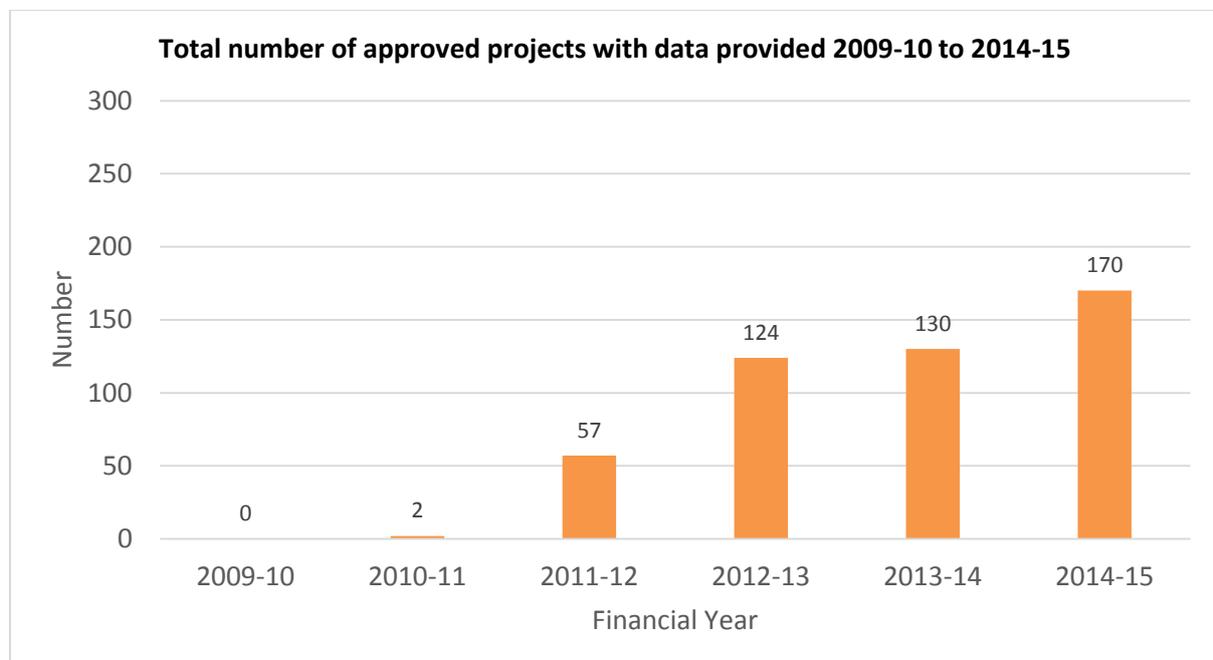
The PHRN conducts coordinated linkage of major statutory and administrative data held by Australian governments. PHRN also has the skills, knowledge and experience to build and operate data linkage infrastructure that meets the requirements of the complex legal and regulatory environment that governs use of health and human services data in Australia.

Linked data is an essential part of Australia’s health and medical research infrastructure. It will underpin innovation over the next 10 years and enable a better understanding of health and disease, improved health and a better health system.

The unique combination of the infrastructure developed by the PHRN, high quality population data collections and recent developments in related infrastructure (e.g. genomics and clinical trials infrastructure) provide new opportunities for Australia to conduct world leading research.

There has been a strong increase in demand for linked data in Australia (see Figure 1 below). This is expected to continue as Australia’s linked data assets are enriched with the inclusion of additional types of data.

Figure 1: Demand for linked data in Australia



In order to meet the needs of Australian researchers over the next ten years Australia will need a national, coordinated data linkage infrastructure which provides facilities and services at all four points of the data linkage pipeline. As recognised in the Issues Paper (Section 5.3.5), the national

data linkage infrastructure will need to be significantly expanded and improved to meet the future needs of the research community.

The expanded national data linkage infrastructure will need to include the following features.

Linkage

A significant expansion of the national data linkage infrastructure is needed to meet research demands. This should cover an increase in the number and type of data collections routinely and nationally linked, development of linkage quality standards and accreditation of linkage units to national/international standards. Australia's future national data linkage infrastructure should include:

- High quality probabilistic linkage (based as far as possible on identifiable data) on a routine/enduring basis for a nationally agreed set of data collections which cover the lifespan (birth to death), molecular to clinical and lifestyle data and all parts of the health system i.e. primary, secondary and tertiary care.
- A national enduring master linkage map/key/file, with continued capacity for ad-hoc linkages (for research cohorts, surveys etc.).
- Additional functionality in response to demand e.g. familial linkage (for genetic association studies etc.) and geocoding.
- National/international linkage quality standards/accreditation and information security accreditation of linkage units.

Research Design

Researchers need a number of services and facilities to assist them to design high quality research projects including:

- Centrally and easily accessible metadata on all routinely linked data collections. The knowledge (metadata) held by individual data custodians about their data collections needs to be systematically captured so that researchers can design high quality projects and better interpret their findings.
- Centrally and easily accessible information on application and approval processes.

Authorisation (Applications and Approvals)

As recognised in the Issues Paper (Section 5.3.5), a more streamlined authorisation environment is needed. This should not only respect individual privacy and confidentiality but also enable research that is demonstrated to be of public benefit. The authorising environment will include:

- Clarification/amendment of specific legislation and regulations where necessary to enable data linkage.
- Minimisation of duplication and overlap of data custodian, human research ethics committee and data linkage unit review.
- Open and transparent application processes which include:
 - Clear criteria for decision making
 - Timelines for review
 - Requirement to provide reasons for decisions
 - An independent appeal process
 - A national online application system.

Access and Analysis

Achieving secure, timely and efficient access to linked data for research purposes will require the expansion of existing facilities and the establishment of new facilities and services including:

- Content data management at jurisdictional level (e.g. jurisdictional repositories) for more effective and efficient data management and release.
- Summary linked datasets for specified high demand topics with simplified data access processes.
- Secure remote access facilities for access to both linked and unlinked personal data and data where there is a risk of re-identification (even if the risk is low).

Collaboration Opportunities

A national, expanded and highly-efficient linkage system will be the glue that enables Australia's health and medical infrastructure to deliver a national co-ordinated approach to the new frontiers of health and medical research.

The PHRN is currently a collaboration of universities, research institutes and Commonwealth and state departments and agencies. This collaborative network approach enables a wide range of linked data to be made available for research and access to linked data to be provided to researchers across Australia on a non-exclusive basis.

An expansion of Australia's data linkage infrastructure is likely to involve collaborations with a range of research organisations and infrastructure providers including the Australian Clinical Trials Association, the Australian Genomics Alliance, Bioplatforms Australia, Therapeutic Innovation Australia, the National Imaging Facility and the Capital Markets CRC.

Summary

Australia needs to build healthy and resilient communities by developing treatments, solutions and preventive strategies to improve physical and mental well-being, improving the efficiency and effectiveness of Australia's health care system and addressing the social, environmental and economic issues that result in family violence, youth disenfranchisement, unemployment and other "wicked problems".

Australian researchers are keen to understand and find solutions to these issues. They will require comprehensive, national and coordinated infrastructure to meet the challenge. As outlined in the Issues Paper, high quality linked data at the unit record level will be an essential tool for researchers across the spectrum of health and human research disciplines from 'omics to clinical trials to health and human services research.

The challenge in making personal or unit record data available for research is the need to balance the right of people to control their personal information with the competing public interest in the research. Australia needs specialised data infrastructure to enable timely access to linked data in a way that minimises risks to individual privacy and ensures the use of this data is legal, ethical and has community support.

The PHRN is a world class, specialised data linkage infrastructure that has the capability to link data from a wide range of sources from genomics to government administrative data. A significant expansion of Australia's data linkage infrastructure is required to meet the needs of the health, medical science and related research communities over the next ten years. This should include:

- An enduring national master linkage map incorporating additional data collections from government, research and private sectors
- National/international accreditation of linkage quality and information security
- Central metadata resources
- Streamlining of the authorising environment
- Expansion of secure data storage and access facilities for linked and unlinked unit record content data.

Understanding Cultures and Communities

Questions 24, 25 and 26: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed? Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond? Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

As explained in the health and medical science section above, Australia has a wide range of linked population-based, health and health-related data. This data includes early childhood, education, justice and families' data, as well as health and medical data. In the short to medium term it is likely that the linked data will include other social data. This data supports research into disadvantaged and other groups. It is directly relevant to the Understanding Cultures and Communities capability.

National Security

Question 27: Are the identified emerging directions and research infrastructure capabilities for National Security right? Are there any missing or additional needed? Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond? Is there anything else that needs to be included or considered in the 2016 Roadmap for the National Security capability area?

Cyber security remains an issue for all data based national research infrastructure capabilities. The PHRN infrastructure holds sensitive data drawn from across the Australian population. Security is therefore of the upmost importance. PHRN has processes in place to manage data security but these will need to evolve in light of emerging data security risks and infrastructure development.

It will be essential to continue to monitor cyber security risks and invest in up to date technology and processes to ensure the security of data within the PHRN. One approach will be to develop methods which do not require the release of name-identifying information, yet allow the linkage of records to a high level of accuracy.

Underpinning Research Infrastructure

Questions 30, 31 and 32: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed? Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond? Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

Underpinning research infrastructure is important to PHRN. AARNet is important for transmission of data across Australia and geospatial data is also important. A number of jurisdictions have extensive geocoding of their statutory and administrative health and human services data collections. Other jurisdictions have prioritised geocoding of their data collections and plan to achieve this in the short to medium term. Geospatial data is important for many health and human services research projects. Large computational facilities are likely to become increasingly important as the scale of health, medical sciences and other human services data increases.

Data for Research and Discoverability

Question 33, 34 and 35: Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed? Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond? Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

Research data management remains an important priority for Australia. Data-rich research infrastructure capabilities have some common needs including the need for high quality, easily accessible metadata. However, there are additional requirements for management of sensitive unit record data including health data drawn from across the Australian population as this data must be dealt with in special privacy preserving ways. The collection, use and disclosure of the data must be lawful and access must be managed in carefully controlled ways. Stakeholders must also be actively engaged to ensure ongoing confidence in and support for use of the data. While PHRN's activities are focused on safe researcher access to linked health and health-related data, we do support a secure data transmission service and a secure access laboratory with a secure data storage function. PHRN has the skills, knowledge and experience to successfully manage sensitive health and health-related data and could provide advice and other support for the storage of health and health-related research data in Australia.