Productivity Commission Submission

The Productivity Commission has been tasked with identifying ways of increasing and improving data linking and availability and the benefits and costs of doing so. This Inquiry provides a real opportunity to improve researchers' access to data and to the resources needed to make greater use of data.

Research Australia is preparing a submission and is seeking your assistance with answering the following questions:

1. **How can better access to data improve Australian health and medical research?**

2. **What are the public and private datasets that can be most useful to health and medical researchers?**

3. **What are the datasets that should be linked together?**

4. **What are the legal impediments to using data for research?**

5. **What other barriers currently exist to making better use of data?**

Health Geography Study Group Submission

Statement to promote access to data for teaching and research.

The Health Geography Study Group (HGSG) of the Institute of Australian Geographers (IAG) provides this statement to emphasise the importance of data access to support policy relevant research. We stress this is required at a scale that improves our understanding of how health and disadvantage are distributed across Australia. The burden of disease is disproportionally represented in low socioeconomic populations and understanding where these populations are located is critical to targeting programs to improve socioeconomic status (SES) and lessen the burden of disease. There are many State and Commonwealth government administrative data that would provide a rich source for researchers to provide the evidence base for assessing and developing policy for environmental and social issues. Many of these data sets are already linked with georeferences that enable integration within geographic information systems and removes the need for complex unique identifiers. All that is required to build large integrated data sets is a spatial engine to link these data. The National Collaborative Research Infrastructure Strategy (NCRIS) funding has built the Australian Urban Intelligence Network (AURIN) at the University of Melbourne. AURIN is built upon a spatial data framework to deliver data and is a facility that could be maintained and further developed as a data portal for government data for the research community. One issue with funded services such as AURIN is the limited nature of the funding that firstly builds and then captures data, but without funding security beyond the funding framework often die once the funding is removed. This will be followed by a period of poorer data access before the next "AURIN" is funded and the wheel reinvented and data access improved. This cycle is both inefficient and wasteful and fails to build on the many lesson learned from the previous iteration of data supplier/service. Governments often speak of open data and data warehouses, but these seldom result in access to data that is structured in a way that it is readily integrated into both submissions for research funding through the Australian Research Council or the National Health and Medical Research Council schemes. AURIN has developed into a data portal that can assist researchers in their funding submissions by ensuring data supply. That is not to suggest that AURIN cannot be improved. It could be improved in a number of ways, firstly with assured research funding, data custodian's could provide data to a service that could licence access from a single point and in the knowledge that it was enduring. Second, the scope of the terms of reference could be extended to serve as a social data archive, storing annual data layers from government to guarantee temporal data supply and provide a data

backup service for the many commonwealth and state agencies. Thirdly, in concert with other national data suppliers work towards national data standards.

For health and medical research, the listing of data sets is vast and exhaustive, as a thorough analysis requires data across systems that is both environmental and social. Some of the main themes for data are listed below:

- Medicare data
- PBBS
- Primary Health care data
- Hospital data
- Linked data
- Births and deaths
- Education
- Locational data for services
    - Health
    - Education
    - Social security
    - Community services
    - Public transport
    - Retail
    - Industry
    - Financial services
- Property data
- Road network
- Business services

It could be argued that many of these data are already available, and that is true, but not nationally consistent in terms of definition or classification or spatially and temporally. Another barrier is the varied cost and licence environment to access these data. While research in a contained study area (such as a city or region) can be managed via access to state or local government for data, it may still require a cost and may not be comparable with similarly themed data from another jurisdiction.

Nationally based research though can often be blocked my issues around consistent data. Nationally important health issues cannot be properly compared due to data differences. One example is the association between health, physical activity and access to open space. There is a large research literature supporting the health benefits and improved levels of physical activity associated with having open space nearby. Many of the studies are based on single cities or regions within cities. To complete a study of this nature in Australia at present is not feasible due to multiple data jurisdictions and varying definitions and standards on data accuracy. Another examples is the limited release of data. Health service location data can be accessed via AURIN, however, the data custodian will not release all of the data even though it would be beneficial for research. Data items collected but not supplied include hours of operation a crucial item if modelling accessibility to services, such as the Cardiac ARIA accessibility work completed in 2010.

This commission is the first step in enabling a broader dialogue about some of the systematic issues that need addressing regarding supply, standards, and definitions and then move onto the identification of key data sets for research and teaching.