



Productivity Commission

Public inquiry: Data Availability and Use

National Health and Medical Research Council Submission

The National Health and Medical Research Council¹ (NHMRC) welcomes the opportunity to respond to the Productivity Commission Inquiry into data availability and use.

NHMRC is Australia's leading body for supporting health and medical research through funding, developing health advice for the Australian community, health professionals and governments and providing guidance on ethical behaviour in the conduct of health and medical research and health care. The functions of NHMRC are legislated under the National Health and Medical Research Council Act (1992)².

NHMRC acknowledges the benefits of making effective use of data, as outlined in the Productivity Commission Issues Paper *Data Availability and Use* (April 2016). In particular, the benefits to be derived from mining, repurposing and reusing data through enhanced data sharing, better data linkage and improved data quality will be significant. NHMRC also acknowledges that the privacy, security, ethical and intellectual property issues associated with data sharing, access and reuse need addressing at a national level.

This submission addresses data generated as a result of NHMRC's public funding of health and medical research, which can include both public and private sector data.

Benefits for health of data sharing, data linkage and better access to data

Data sharing, linkage and access to health and medical research across government and academia can have a direct impact on health care and health outcomes, through more informed policy making and evidence-based clinical practice.

¹ <https://www.nhmrc.gov.au/>

² <https://www.legislation.gov.au/Details/C2014C00364>

For example, Cole *et al.* (2014) described a data sharing network across five states in the United States of America and nine primary care organisations where better use of resources and quality of care was achieved. The Western Australian Pregnancy Cohort (Raine) Study³ is another cohort study with powerful data linkage capabilities that examines the role of early life events on later health. Many successful outcomes have been derived from this study that encompasses all areas of health from cardiovascular diseases to mental health and gastroenterology. The benefits that could be derived from sharing health data, creating strong linkages between datasets and creating better access pathways to such data, are well documented in many other health areas including cancer genotyping (Bailey *et al.* 2016), diagnostics for kidney disease (Papadopoulos *et al.* 2016) and treatment regime for cancer (Sharma *et al.* 2016).

Sharing of data will encourage better transparency in experimental methods, assist in improving quality of data through increased scrutiny, reduce duplication and hence wastage of limited resources and allow repurposing of data to occur (Walport & Brest 2011). Linkage of data between datasets allows new discoveries to be made in a highly cost-effective way, and by combining datasets the power to answer questions about health is increased—effectively leveraging the Commonwealth's investment in health and medical research. Gertig *et al.* (2013) showed through data linkage that human papillomavirus vaccinations in Australia significantly reduced cervical abnormalities while Mathews *et al.* (2013) successfully investigated cancer incidence rates for people exposed to CT scans.

However, many challenges exist, including ethics and privacy concerns associated with linkage between data sets created for differing purposes. Australia's investment in the Population Health Research Network⁴ (PHRN) is an example of a network that has made significant progress in addressing issues such as these. The PHRN has demonstrated that data access pathways under a strong governance framework can provide secure, reliable data access for researchers, clinicians and policy makers while respecting the expectations of individuals and the community.

Ultimately better use of data will not only improve health care, it will also inform health economic considerations for health system designs and strategies (Lorgelly *et al.* 2016).

NHMRC data and policies

NHMRC recognises that there is a need to balance the protection of privacy and the facilitation of access to data for research purposes. In relation to ethics, NHMRC provides guidance through guidelines in the *National Statement on Ethical Conduct in Human*

³ <http://www.rainestudy.org.au/>

⁴ <http://www.phrn.org.au/>

*Research 2007*⁵, the Guidelines under Section 95 of the Privacy Act 1988, 2014⁶ and the Guidelines approved under Section 95A of the Privacy Act 1988, 2014.

The work of NHMRC is associated with two major categories of data. The first is administrative data (hence, public sector data) which includes all the data required to manage and report on the research grants made by NHMRC. Reports on funding decisions and other policy relevant outcomes of the grant programs are made publicly available. The NHMRC has commenced development of a data strategy along with a process of streamlining its Information and Communications Technology operations. It is envisaged that NHMRC will be able to provide a new virtual office environment and data visualisation tools that will allow more open access to NHMRC funding data for analysis. This will enable more work to be undertaken on identifying the return on Government investment. NHMRC's data warehouse and reporting infrastructure may also create efficiencies with grants processes including the ability for other health and medical research funding programs to benefit from this capability.

The second and far larger category is data resulting from research funded by NHMRC. NHMRC has developed policies to require or encourage grant holders to share various outputs of research funded by NHMRC as widely and as early as possible. This approach is to ensure that maximum benefits are derived from the investment in research. NHMRC has approached this through three pathways: sharing, access/use and linking researchers:

A: Sharing

NHMRC Open Access Policy for Publications⁷

NHMRC requires that any publication arising from NHMRC supported research must be deposited into an open access institutional repository and/or made available in another open access format within a twelve month period from the date of publication.

NHMRC Statement on Data Sharing⁸

NHMRC encourages research institutions to share data and provide access to data and other research outputs (metadata, analysis code, study protocols, study materials and other collected data) that arise from NHMRC supported research. This statement aligns with researchers' responsibilities under the Australian Code for the Responsible Conduct of Research (2007)⁹, which provides advice on the storage, management and privacy of

⁵ <https://www.nhmrc.gov.au/guidelines-publications/e72>

⁶ <https://www.nhmrc.gov.au/guidelines-publications/pr1>

⁷ <https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy>

⁸ <https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing>

⁹ <https://www.nhmrc.gov.au/guidelines-publications/r39>

research data (section 2.5-2.7) and states: “Research data should be made available for use by the other researchers unless this is prevented by ethical, privacy or confidentiality matters.” It is also aligned with the National Statement on Ethical Conduct in Human Research (2007)¹⁰.

B: Access

NHMRC Principles for Accessing and Using Publicly Funded Data for Health Research¹¹

The NHMRC *Principles for accessing and using publicly funded data for health research* are intended to facilitate health and medical academic and researcher access to public administrative data sets (such as national mortality data and national hospital data). This set of principles ensures all parties have a common set of priorities for the use of data, encourages communication between researchers and data custodians so they can better understand each other’s roles and responsibilities, and helps to optimise data use for the benefit of all Australians. The document was developed in consultation with the Consumer Health Forum of Australia, the Australian Government Department of Health, the Australian Institute of Health and Welfare, the Australian Bureau of Statistics, the Australian Government Department of Human Services, the Australian Electoral Commission, the Australian Institute of Aboriginal and Torres Strait Islander Studies and Universities Australia.

C: Linking researchers

NHMRC and ARC Statement on Open Researcher and Contributor ID¹²

NHMRC and ARC encourage all researchers to use ORCID identifiers, a global system with capabilities to disambiguate researchers and research outputs. It also enables the linking and reuse of high quality, persistent data (e.g. publications, grants), to minimise administrative burden, minimise reporting burden and improve research output connectivity.

Key areas of importance for development in data availability and use

A: National data infrastructure

The NHMRC encourages academic and research sectors to disseminate research findings as broadly as possible. Linkage infrastructure is also in place across Australia through initiatives such as the PHRN (mentioned previously) that facilitate academic and researcher access to public administrative data sets. Another facility currently being developed is

¹⁰ <https://www.nhmrc.gov.au/guidelines-publications/e72>

¹¹ <https://www.nhmrc.gov.au/principles-accessing-and-using-publicly-funded-data-health-research>

¹² <https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-and-arc-statement-open-researcher-and-contributor-id-orcid>

med.data.edu.au¹³ that combines a well governed data storage and access facility with high performance computing services.

While academic research data is not explicitly in scope for the current Productivity Commission inquiry, data sets created by health and medical research academics and research facilities are valuable resources. Such data can only be made open access and used effectively for the long term if there are resources available that allow that data to be managed and curated properly beyond the period of funding for each research program. This could be facilitated through either University level infrastructure or through national data storage and data sharing infrastructure (e.g. Secure Unified Research Environment¹⁴). Such infrastructure would need to include metadata management and support. In addition, researchers would need to be aware of national metadata standards and be properly resourced to ensure they could comply with such standards.

European data infrastructure models could be investigated as an example on how to curate academic and research data from the Australian research fields. Euro Cloud¹⁵, EUDAT¹⁶ and other more specific data cloud initiatives such as Jetstream¹⁷ in the USA that provide data storage, linkage and high performance computing could be further explored for their applicability to the Australian data landscape.

In addition to curated, open access data sets and capability to link these with public data sets (e.g. PHRN, Australian Bureau of Statistics, Australian Institute of Health and Welfare), it is also important that the health and medical sector has capacity to make use of the data. An investment may be required to train more data scientists to ensure Australia has the capacity to make use of the existing data.

Secure funding streams will enhance Australia's data infrastructure. There are currently few pathways available to guide individuals, institutions and organisations as to how this infrastructure and access to it can be funded. If the system relies on a user pays model there will be many interested parties that will be disadvantaged because of lack of funds.

Any form of data infrastructure should also be designed and managed with appropriate data governance to ensure that data usage and sharing are ethical and privacy concerns are addressed.

¹³ <http://med.data.edu.au/>

¹⁴ <https://www.saxinstitute.org.au/our-work/sure/>

¹⁵ <https://www.eurocloud.org/>

¹⁶ <https://eudat.eu/>

¹⁷ <http://jetstream-cloud.org/>

B: Public sector data

A single, clear set of protocols for all stakeholders to request access to public sector data (irrespective of which party currently holds the data) would be valuable to ensure the efficient use and linking of data. Researchers have also provided feedback to NHMRC about their concerns that it is difficult to ensure timely access to data, and ensure it is affordable. Hence the protocols would need to be implemented with appropriate resources to support efficient access and appropriate analyses of the data. A clear and consistent pathway for data access and use of public sector data will ensure that research is expedited, resources are not unnecessarily expended in obtaining the data, better transparency is achieved in relation to privacy and ethical aspects of the data used and allow better planning to occur. Funders at all levels (both public and private sector), interested in investing in research involving data linkage, will also be reassured that the project will be achievable within reasonable timeframes and budgets.

C: Private sector data

There are legitimate and significant benefits for researchers and clinicians in accessing private sector health data for research (for example where including whole of population data reduces bias in analysis). Excluding privately held health data from analysis limits conclusions that can be made about health service process improvement and inhibits health system wide outcomes reporting. To report on the efficiencies of the health sector in Australia it is vital that all health service data is included in analysis and reporting. This requires that private sector health data is made available in the same disaggregation as public sector health data. This will require cooperation from the private sector and a policy framework that takes account of ethical and legal considerations, cost arrangements, contractual expectations, timeframes for the processes and any technical considerations.

Summary

NHMRC supports efficient and effective use of data at all levels, provided that health and medical data is protected by appropriate governance arrangements to ensure ethical use and individual privacy concerns are addressed. This will maximise benefits to be derived from available data, and ensure that duplication of effort is reduced and wastage of resources is minimised.

Data infrastructure, appropriately resourced, will be needed to achieve this vision. Other important building blocks are protocols to ensure access to public sector data is consistent, timely and affordable; and inclusion of private sector health data to allow truly whole-of-sector analysis.

References

Bailey P, Chang DK, Nones K, *et al.* 2016. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531(7592):47-52. doi: 10.1038/nature16965.

Cole AM, Stephens KA, Keppel GA, *et al.* 2014. Implementation of a Health Data-Sharing Infrastructure Across Diverse Primary Care Organizations. *J Ambul Care Manage.* 37(2):164–170. doi: 10.1097/JAC.0000000000000029.

Gertig DM, Brotherton JML, Budd AC, *et al.* 2013. Impact of a population-based HPV vaccination program on cervical abnormalities: a data linkage study. *BMC Medicine* 11:227. doi: 10.1186/1741-7015-11-227.

Lorgelly PK, Doble B, Knott RJ *et al.* 2016. Realising the Value of Linked Data to Health Economic Analyses of Cancer Care: A Case Study of Cancer 2015. *PharmacoEconomics* 34: 139. doi:10.1007/s40273-015-0343-2.

Mathews JD, Forsythe AV, Brady Z, *et al.* 2013. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ* 346 :f2360. doi: <http://dx.doi.org/10.1136/bmj.f2360>.

Papadopoulos 1, Krochmal M, Cisek K, *et al.* 2016. Omics databases on kidney disease: where they can be found and how to benefit from them. *Clin Kidney J.*: 9(3):343-52. doi: 10.1093/ckj/sfv155.

Sharma RA, Plummer R, Stock JK, *et al.* 2016. Clinical development of new drug–radiotherapy combinations. *Nature Reviews Clinical Oncology* (Published online 01 June 2016) doi:10.1038/nrclinonc.2016.79.

Walport M, Brest P. 2011. Sharing research data to improve public health. *Lancet* 377 (9765):537 – 539.