

Review of Health and Medical Research in Australia.

ANDS' comments on selected terms of reference.

ANDS, the Australian National Data Service, welcomes the opportunity to comment on three of the terms of reference of this timely review. These are 3, 5 and 11, as below:

3. (...) opportunities for more efficient use, administration and monitoring of investments and the health and economic returns; including relevant comparisons internationally,
5. Likely future developments in health and medical research, both in Australia and internationally,
11. Ways in which the Commonwealth's e-health reforms can be leveraged to improve research and translation opportunities, including the availability, linkage and quality of data.

The nature of research is changing in many ways, some of which relate to developments enabled by improvements to information and communications technologies. These in turn relate to a focus on data, its capacity to be aggregated for meta-analysis and its capacity to be shared with co-researchers locally and internationally during the course of research and more broadly at a later date.

In health and medical research, the proliferation of instruments such as gene sequencers or MRIs has created unprecedented new volumes of data for analysis. At the same time the emergence of networked information about population health and clinical observations has enabled new approaches to medical and health research such as data linking, data mining, and meta-surveys. These developments combined are introducing a new paradigm of data-driven research to complement the existing empirical, theoretical, and computational approaches. The result is not just more efficient or effective research, but new kinds of research that could never have been previously contemplated.

Given the centrality of data to future developments in health and medical research, this submission from ANDS highlights the new policy, infrastructure, and operational needs of this data-driven science. Having said that, the infrastructure etc required to support this kind of science is equally applicable and useful to the more traditional forms of research, and offers significant benefits in terms of efficiency and cost.

This submission includes:

1. an overview of ANDS
2. an overview of international trends in data sharing for health and medical data
3. the needs of data driven science and funding implications

1. ANDS

ANDS (www.ands.org.au) was established with a charter to create the Australian Research Data Commons. The Australian Research Data Commons supports the discovery and access to research data held in Australian universities, in publicly-funded research agencies and in government organisations for use by researchers. Collectively, our domain might be referred to as ‘public sector information’ recognising that this is broader than just government organisations and that governments fund much of the activity of the research sector in universities and elsewhere.

ANDS supports the construction of a range of ICT utilities to capitalise on and ensure greater use and re-use of existing data resources, as well as better management of new data generated in Australian research in all disciplines.

One way to access the contents of the Australian Research Data Commons is through [Research Data Australia](#)¹. This is a set of interlinked, user-friendly web pages that display rich descriptive information contributed by universities, research organisations and public sector organisations throughout Australia. While most searchers will seek information about data through web search engines, Research Data Australia is designed to add value by offering many ways of searching. ANDS does not collect or hold the data themselves—rather it collects the metadata which describes data created in the course of research. Data of interest to researchers in health and medicine is definitely also in scope for Research Data Australia.

ANDS has a remit to ensure that data is shared ethically and legally. To this end it has developed a *Guide to Ethics, Consent and Data Sharing*² which provides practical guidance for both researchers and Human Research Ethics Committees.

2. International perspectives on data dissemination and sharing

There is strong interest throughout the world in ensuring that data is exploited in order to improve research outcomes and ultimately provide benefit to humanity. The trend towards data exploitation and sharing can be seen in the policies of governments and research funding bodies, such as the National Science Foundation, National Institutes of Health, UK Research Councils, Cancer Research UK and the European Union.³

For example, the Medical Research Council in the UK states:

Every year, the MRC invests around £700 million of public money in research, the primary output of which is data. We want to maximise the research opportunities that such a diversity, richness and quantity of data provides. One of the best ways of achieving this is to ensure that

¹ Research Data Australia can be viewed at <http://services.ands.org.au/home/orca/rda/>

² <http://ands.org.au/guides/ethics-working-level.html>

³ For an overview of international policies see: <http://biosharing.org/?q=policies>

data are properly preserved for sharing and informed use beyond the originating research teams.⁴

Fundamental to the idea of data sharing is the need for data to be carefully managed and curated, in both the short and long term. In the case of health and medical data, it is important that ethical issues be addressed early on, again referring back to the guide we have just created. In both the UK and US, some funding bodies are now requiring that data management plans are included with funding applications. While the requirements are referred to as ‘data management plans’, in practice the focus is on the dissemination of research outputs, covering both publications and data, and how this will be achieved.

In the UK, where funding bodies fund both research projects and data centres, funder requirements can be more specific about data management as well as dissemination of research outputs. In Australia, however, research funding bodies focus necessarily on dissemination of data outputs, since data centres and data management are largely resourced by institutions and national or state-based research infrastructure.

3. Strategic needs for data-driven medical and health research in Australia

A ten year plan for health and medical research will need to address the specifically new needs of data-driven research⁵, most of which were not prominent in previous decades. These strategic needs are dealt with here in three categories:

- data management infrastructure and funding implications
- data sharing policy
- data operational requirements

A. Data management infrastructure

Data-driven research relies on the existence of a managed corpus of data with which to work. Data management infrastructure is therefore required both for research data (e.g. the output of NHMRC research) and for administrative or public sector data.

As noted above, Australian research funders provide funding for research, but, unlike their UK counterparts, do not fund data centres where data can be stored, curated and made available, under strict access provisions where these are required. This has led to a serious emerging problem in

⁴ <http://science.cancerresearchuk.org/funding/terms-conditions/funding-policies/policy-data-sharing/>

⁵ repeating an earlier point: data-driven research requires this infrastructure; more traditional research forms can benefit from it in terms of cost, access and efficiency.

Australian research infrastructure provision, and one which is not confined to medical research. It is demonstrably the case that the emerging trends in data-intensive research (e.g. eResearch)—brought about largely by wholesale changes in ICT and related technologies—has created a vacuum in this country in terms of long term data storage, curation and discovery infrastructure.

Provision of data storage is not sufficient to meet the burgeoning needs of medical and health research. While there are some excellent data storage facilities available, these have not been set up with the facility to manage, curate, preserve and provide mediated access in the longer term. An example of a large capacity storage facility is the the Research Data Storage Infrastructure (RDSI) Project, which is an initiative of the Department of Industry, Innovation, Science, Research and Tertiary Education, and funded from the Education Investment Fund under the Super Science (Future Industries) initiative.

Conceptually, at least, a combination of the discovery capability of Research Data Australia infrastructure (described earlier) and RDSI would take care of discovery and storage components of medical and health research data sets, but a ten year plan for health and medical research strategy will require a significant investment from the Department of Health and Aging, akin to the model well established within DIISRTE, to ensure a long term strategy for innovation; importantly, the DIISRTE model includes data management infrastructure, data sharing policy, and data operational requirements⁶. An relevant example in the health sector is the Population Health Research Network (<http://www.phrn.org.au/>).

A strategic ten year plan for the health and medical data will contain a plan for the data management needs of research and public sector organisations trying to enable data-driven research by making existing data available for further analysis.

Key point 1: creating the appropriate data management infrastructure will require investment by the Health and Aging portfolio, perhaps leveraging off, coordinating with, and complementing with the approach of DIISRTE in relation to innovation.

B. Policy framework:

An appropriate policy framework for data-driven health and medical research will be based on a default setting for openness and data re-use, tempered when necessary by valid ethical and privacy constraints. Again such a comprehensive policy framework will span both public sector data created through the health care system as well as research data created through the research system.

An appropriate policy framework needs to be developed, such that, for example, the NHMRC application and reporting templates support data re-use and sharing; policy principles need to be developed for mediating access to data in the data store; frameworks need to be developed which encourage the appropriate use of data licenses.

⁶ <http://ncris.innovation.gov.au/Pages/default.aspx> <http://www.innovation.gov.au/Science/ResearchInfrastructure/Pages/default.aspx>

In Australia, both the ARC and the NHMRC are developing and implementing policies related to data sharing in accordance with the directions set out in the *Australian Code for the Responsible Conduct of Research*. At this stage, neither have data planning requirements as specific as those of overseas funding agencies, although this may change in the short to medium term.

Equally in the public sector, consent and privacy policies, for example, need by default to enable data-driven research whilst ensuring sound ethical constraints.

Key point 2: A ten year plan that supports data-driven health and medical research will need to develop a comprehensive policy framework to enable data re-use.

C. Data operational arrangements

There are a number of operational arrangements that data-driven research requires, two of which are highlighted here: re-use permissions and tracking.

Re-use permission: an effective and practical data licensing scheme needs to be used to ensure clarity in terms of re-use permissions, terms, and conditions. Prospective re-users need to know exactly what they can and cannot do with the data. Equally, the creator of the data set needs appropriate citation and attribution (such a licensing framework exists in Australia⁷).

Tracking: effective tracking of data re-use needs to be employed (e.g. via Digital Object Identifiers) so that all relevant parties can assess the effectiveness of their policies and practices. DOIs, for example, would allow NHMRC to measure the use and re-use of different classes of medical and health research data sets; the creators of the data sets will receive data citations, which, increasingly in the international arena, are increasingly being recognised as legitimate research output⁸ (infrastructure around DOIs exists in Australia⁹).

Key point 3: Data management infrastructure and policy frameworks need to be supported by licensing and tracking tools to ensure optimal outcomes and return on investment.

Summary

Failure to address the problem of long term medical and health research data management infrastructure in Australia will certainly decrease the return on investment into medical research. It will also deny many current and future researchers, the opportunity to take advantage of the emerging trends around data-intensive research.

⁷ <http://www.andc.org.au/publishing/licensing.html>

⁸ <http://www.ijdc.net/index.php/ijdc/article/view/208>

⁹ <http://www.andc.org.au/guides/doi.html>

For an innovative country like Australia, this would be an untenable outcome.

DIISRTE has created this infrastructure for the innovation sector; this is now largely operational in Australia. The Health and Aging portfolio should follow suit to ensure the medical and health sector remain at the cutting innovative edge of medical and health research, embracing the new data-intensive research paradigm. This will comprise three areas:

- data management infrastructure, which will require investment by the Health and Aging portfolio
- a comprehensive policy framework to enable data re-use (which is)
- supported by licensing and tracking tools to ensure optimal outcomes and return on investment.

ANDS would welcome the opportunity to discuss these issues in more detail during the enquiry. ANDS is in a unique position in terms of research data infrastructure in Australia. Whilst we do not collect or hold these data sets, we are tasked with creating the infrastructure around their discovery. As such, we are working with many agencies in the research and innovation sector, including in the medical and health sector. We believe that there is a significant opportunity for the Review of Health and Medical Research in Australia to address the emerging shortcomings in medical and health research data infrastructure, as outlined in this submission. Our experience to date, suggests that there are already cost-effective ways to address this issue, essentially by building on existing infrastructure like RDA, as well as using existing licensing and tracking infrastructure.

ANDS is also well advanced in terms of policy frameworks, DOIs, ethics and so on, and has created a range of guides and reports on these issues. We would be happy to provide these to the Review.

The contact officer for this submission is Dr Greg Laughlin on (02) 6125 1175 or gregory.laughlin@ands.org.au.