

DATA AVAILABILITY AND USE
PRODUCTIVITY COMMISSION ISSUES PAPER – APRIL 2016
QUESTIONS

2. BENEFITS OF INCREASING DATA AVAILABILITY AND USE (STARTS PAGE 9)

Page 14

QUESTIONS ON HIGH VALUE PUBLIC SECTOR DATA

What public sector datasets should be considered high-value data to the: business sector; research sector; academics; or the broader community?

- i) Longitudinal capacity: ability to monitor and hence predict societal trends (eg in adverse outcomes, or protective factors or lifestyle choices) which may inform policy and enable evaluation. Also useful for service planning or for products of value to community; academics and researchers also interested to enable analyses of causes, risk and preventive characteristics.
- ii) Geographic and sub-population variations to enable better provision of services and products in areas of highest need (GIS coded data is very high value for this – see note below) – eg: in WA the police wanted to map whether their services were appropriate for “hotspots” of crime, domestic violence, child abuse and risk factors for these (social, geographic, sub-populations, unemployment, school truancy, renting public housing etc)
- iii) Data sets that provide information on:
 - a) what works (how cost effective? Better than doing nothing or something else?);
 - b) what doesn't (costly but useless);
 - c) what's harmful and
 - d) population characteristics of what works, what doesn't, what's harmful etc and why – this is enabled by record linkage of course.

Note on GIS coding of data: we have GIS coded all data sets in our WA linked data sets; it enables linkage of outcomes\events to exposures such as environmental teratogens, carcinogens, water, soil and air pollution, climate data (temperature, humidity, rainfall etc), aircraft paths, traffic\road use, green space, schools, shopping centres, alcohol outlets and to services such as medical centres, hospitals, schools, police stations etc

What characteristics define high-value datasets?

The characteristics include:

- i) Available on the total population; this means that there is no bias in who is selected or followed up or interviewed or who responds to a questionnaire or a survey. Total population data means that no-one is left out (eg youth, elderly, mentally ill, Aboriginal,

rural locations etc). The power of public sector data is that all people who are registered (eg Medicare, birth certificates), who use a service (hospital, school, employment agency, training course, prison) are included and data recorded. The reason why there is so much variability in many outcomes from research is that the analyses from observational studies with variable participation rates result in different associations being observed.

- ii) Complete ascertainment.
- iii) Accurate information on a broad range of important factors.
- iv) Clear definitions of personal details, socio-economic factors etc to enable analyses which can control for likely confounding or mediating factors.
- v) Accurate identifiers to enable individual data to be linked.
- vi) Unit record data are essential if specific policies and practices are to be evaluated; unless the individual “exposure” to the service is recorded there is no way that the outcome can be attributed to the service to be evaluated. Just comparing average or distributions of outcomes before and after a service does not constitute good evaluation as many other confounding and mediating factors may have also changed and be responsible for the change in the outcomes observed, including the risk and protective factor profiles of those receiving it.
- vii) Health, Education, Training, Employment, Family, Living standards, Housing and Environmental data are all high-value because of their importance to the problems facing our societies (see below).

What benefits would the community derive from increasing the availability and use of public sector data?

There are many benefits:

- i) Modern societies are facing a range of expensive, complex problems (“wicked” problems) which demand data across a range of agencies to monitor, identify risk and protective factors and seek evidence based solutions. Such problems come from health (obesity, mental ill health, chronic diseases, substance abuse, inequalities), education (learning disorders, disengaged youth, inequalities), child neglect, and maltreatment, juvenile crime and incarceration, environmental degradation, climate change, overconsumption, and economic and other inequalities (including the gaps in most outcomes between Aboriginal and non-Aboriginal people in Australia). In order to monitor, understand the causes of, and develop preventive strategies for, these wicked problems, we need data from many agencies linked together. For example social determinants of health outcomes include educational, economic, geographic and racial factors some of which are protective and some risks. If we really want to improve these outcomes, it is likely that the solutions may come from multiple agencies; siloed thinking and policies do not work in terms of prevention. Hence the first huge benefit to the community from joining up public sector data is to provide an understanding of the causal pathways and the effective ways to prevent these problems. Many of the health, educational, mental health, child maltreatment and crime outcomes arise from similar sets of risk factors and hence there is power in bringing these data together. For example if we implement effective practices to improve health outcomes, then we will also improve education and reduce child maltreatment.
- ii) Obtaining and analysing public sector data can enable a proper evaluation of whether services are of value, are cost effective, and are useless or even harmful. There are very few

services provided by any government which have been properly evaluated as to whether they are effective, in whom and for whom they are not. Do we have any idea why so many services fail Aboriginal people? (Our WA research suggests that most services developed for Aboriginal people ignore the causal pathways, the contexts in which they are living and do not involve Aboriginal communities in the decision making; when these are acknowledged, people do use the services and get value from them). We have been proposing a linkage between the Pharmaceutical Benefits Scheme and all health outcomes since the 2000s which would be a much more effective method than current doctor reporting to develop a rapid adverse drug effects capacity and save lives and money (Stanley FJ, Meslin EM. Australia needs a better system for health care evaluation. Medical Journal of Australia <https://www.mja.com.au/journal/2007/186/5/australia-needs-better-system-health-care-evaluation>)

- iii) Economic analyses of public sector data would demonstrate whether investing in preventive strategies were more worthwhile than an excessive focus on end stage problems, not just in health but education, child maltreatment, environmental and climate change issues. 4, People in the community would be informed about services and make better choices to improve their health, development, education and lifestyles and to encourage behaviour change to improve the environment and the communities in which we live.

QUESTIONS ON COLLECTION AND RELEASE OF PUBLIC SECTOR DATA

What are the main factors currently stopping government agencies from making their data available?

I can only guess at these but they may include:

- i) Lack of understanding of data, analysis, statistical competencies – some agencies have almost a fear of data and hence do not want it to be used or analysed.
- ii) A fear that the data will be used to demonstrate that the services they provide are ineffective or not cost effective or do not help some population groups etc.
- iii) Lack of time to use the data even though analysing their data may allow for major efficiencies.
- iv) Fear of privacy being breached and people being harmed by disclosure of personal information.
- v) Expense of obtaining, using, analysing the data.

How could governments use their own data collections more efficiently and effectively?

In several ways including:

Partnering with academic and other groups and agencies whose capacity for analysis, linkage, interpretation etc is very strong (epidemiologists, statisticians, specialist scientists, economists and other social scientists). There are several models for these partnerships:

- i) opening up data sets with linkage capabilities to academics and researchers in centres, institutes and universities (this is best done as an academic:bureaucratic collaboration); they would then have to get their own funding to analyse which has a long time lag as NHMRC and ARC grants take about 18 months to be funded from planning and writing the grants (success rates now below 20%!).

- ii) Jointly funded state/territory and federal collaborations (like a tender – using our WA or CHRL model in NSW or Manitoba Model in Canada) which engage analytical capacity from researchers in academic centres to conduct a program of work demanded by the bureaucracy.
- iii) Joint ABS/AIHW collaborations with Government departments with or without academics from the universities and institutions.
- iv) Build in house analytical capacity which will take time but is desirable even if governments want to participate in the other options listed here.
- v) Setting up a national Institute of Population Statistics with a hub and spoke model to harness all the state and territory expertise coordinated from Canberra (involving ABS, AIHW, etc). This model could be based on a revamped and expanded AIHW as there may be a reluctance to establish another Federal agency.

NB It is important to note that the main reason for WA having such a long standing and successful system for data linkage and analysis is that the whole process was (and still is) driven by passionate users of data not by the public servants. It is facilitated by having computerised data over a long period of time but the main reason for the quality data sets, the linkage capacity, the sophisticated analyses and the extensive nature of data available is that the researchers realised the enormous value of having such an environment and were prepared to fight for it and show how valuable it is.

Should the collection, sharing and release of public sector data be standardised? What would be the benefits and costs of standardising? What would standards that are ‘fit for purpose’ look like?

In an ideal world it would be of considerable benefit to standardise the collection of public sector data and to have all the variables which are collected on individuals or other aspects of data collected in the same way, using the same descriptors etc. However this will take some considerable time and to ensure that we can benefit by using existing data sets that are disparate now, then we need to develop effective methods of linking these data which enable comparisons to be made when there are different ways of recording and measuring between agencies. There is expertise around such methodologies in many state based groups in Australia. With respect to sharing and release, these certainly can be standardised in ways that serve different purposes. For example giving data to the public for better informed decision making would have certain standards which could differ from those that are for detailed and complex analyses of causes or services.

What criteria and decision-making tools do government agencies use to decide which public sector data to make publicly available and how much processing to undertake before it is released?

I don't know and think that these probably are ad-hoc and not clear within or between agencies and this is certainly an area which the Commission needs to address.

What specific government initiatives (whether Australian Government, state, territory or local government, or overseas jurisdictions) have been particularly effective in improving data access and use?

- i) WA and NSW Governments initiatives in population data and data linkage, much of it around health but now also including a wider group of agencies such as education, child protection and disability.

- ii) National Collaborative Research Infrastructure grant to establish the Population Health Research Network (now also broader than health data) which was set up in WA to take the best methodologies for data linkage to all states and territories. This means that now all states and territories have data linkage capacity which can be harnessed for national use. There is also value in comparing different policies and other data between states.
- iii) Countries that have invested heavily in these capacities and brought in models of governance that are worth us looking at include New Zealand, Canada, Scandinavia, Scotland, Wales and less effectively but with huge investment, England. I know less about other European countries. Asia and the Pacific are way behind even in getting basic data on health and other outcomes. ?China. USA has some models of longitudinal data and very good registers of disease but their private systems preclude total population data with linkage being done extensively. Very good analyses are done using HMO data in USA such as the large Kaiser Permanente data sets out of California and the Mayo Clinics in Rochester. I am not aware of linkages to other data than health in the USA.

QUESTIONS ON DATA LINKAGE

Which datasets, if linked or coordinated across public sector agencies, would be of high value to the community, and how would they be used?

WA model of the Developmental Pathways Project – information already sent – which links data from 13 government agencies and engages closely with agency bureaucrats in the studies done and how the data are used. The leadership of the project was shared between a leading academic in the Telethon Kids Institute (F. Stanley) and the Head of Treasury (Tim Marney who is now head of The Mental health Commission). CHRL in NSW connected with the Sax Institute is another model.

Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?

There are many problems within governments which create barriers to linking data sets:

- i) Unjustified fear of breaches of privacy and confidential data being used inappropriately, in spite of evidence of this being extremely low; in fact there has not been one breach in 35 years of data linkage in WA! Some ethics committees and guidelines classify data linkage risk as negligible. Trust in the system of privacy protected record linkage and analysis as in the WA model must be demonstrated to the bureaucrats and custodians to overcome these fears.
- ii) Lack of appreciation of not needing consent for access to population data or its linkage (see Stanley, FJ “Where Not seeking Consent is Best Practice” Royal Statistical Society UK. 2010)-document sent.
- iii) Data Custodians develop rules and regulations in ignorance of researchers’ needs and capacities, and usually without consultation which results in guidelines which are restrictive and not conducive to good research and data practices.
- iv) The current policies of researchers having to apply for new linked data afresh for all new projects instead of being able to keep linked data sets for multiple analyses and studies; such data sets which take considerable time and effort to produce could be cost-effectively updated and multiple new analyses allowed. This is particularly annoying when the initial analyses give rise to a series of new questions which could be quickly addressed by further interrogation of the data set (which has to be destroyed).

- v) Reluctance to let data out to researchers in spite of NHMRC and other guidelines protecting the use and avoiding the mis-use of such data. Relationships between custodians, bureaucrats and researchers need to be strengthened and trust built if data are going to be made really useful in Australia.
- vi) Lack of understanding that unit record data are required to properly evaluate any services.

How can Australia's government agencies improve their sharing and linking of public sector data? What lessons or examples from overseas should be considered?

See above

QUESTIONS ON HIGH VALUE PRIVATE SECTOR DATA

What private sector datasets should be considered high-value data to: public policy; researchers and academics; other private sector entities; or the broader community?

In each case cited, what characteristics define such datasets?

What would be the public policy rationale for any associated government intervention?

What benefits would the community derive from increasing the availability and use of private sector data?

Whilst I have little expertise or experience in these areas there is a considerable opportunity to obtain data from the private health insurance data bases which contain information on diseases, diagnostic tests, treatments, procedures and out of hospital experiences of patients which are not available in the agency data bases which are currently used by researchers and others. These opportunities should be canvassed by the Commission, as the same methods of protecting privacy and record linkage could be used as with public sector data. Data from the corporate sector could also be useful for understanding better ways to improve health and wellbeing –eg supermarket shopping habits.

PAGE 18

QUESTIONS ON ACCESS TO PRIVATE SECTOR DATA

I have no comments on these questions.

Page 21

QUESTIONS ON CONSUMER ACCESS TO, AND CONTROL OVER, DATA

What impediments currently restrict consumers' access to and use of public and private sector data about themselves? Is there scope to streamline individuals' access to such data and, if there is, how should this be achieved?

Are regulatory solutions of value in giving consumers more access to and control over their own data?

Are there other ways to encourage greater cultural acceptance amongst businesses of consumer access to data about them?

What role do third party intermediaries currently play in assisting consumers to access and use data about themselves? What barriers impede the availability (and take-up) of services offered by third party intermediaries?

What datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, would provide high value to consumers in helping them make informed decisions? What criteria should be used to identify such datasets? What, if any, barriers are impeding consumers' access to, and use of, such data?

It is very important to get good consumer input in to these questions. I suggest Anne McKenzie (UWA and Telethon Kids Institute) who is a world expert on consumer participation in medical research and record linkage and Karen Carey who is nationally regarded as very strong in these areas.

There are two issues here – the set of questions you outline above which relate to individuals having access to their data and how to do that and secondly (which you do not really address) is how consumers view the collection, linkage and analysis of their data without consent for the greater good. There is now considerable consumer support for this work and a sense of annoyance and frustration that such data are not being used to avoid harm and improve services. These views are more widespread than the negative comments from small numbers of ex-privacy commissioners who appear to only have the views of privacy concerns without acknowledging the enormous benefits that accrue from collecting, linking and analysing the data (with privacy protected).

We, as researchers and users (including the policy people and bureaucrats) must appreciate the huge responsibility we have to the consumers whose data we are using, to ensure that we have their trust. Hence we must be transparent as to what data are collected, how managed and accessed, how linked and analysed and provide really good examples of the public good of such work. Such good relationships have been built up over 30 years in WA; those with serious diseases and conditions tend to be very supportive of any research and a large proportion of those who are healthy now appreciate the value of such work to ensure that they have services that work and that we are interested in keeping them healthy.

3. MANAGING THE COSTS

Page 22

QUESTIONS ON RESOURCE COSTS OF ACCESS

How should the costs associated with making more public sector data widely available be funded?

To what extent are data-related resources in agencies being directed towards dealing with data management and access issues versus data analysis and use?

What pricing principles should be applied to different datasets? What role should price signals play in the provision of public sector data?

Is availability of skilled labour an issue in areas such as data science or other data-specific occupations? Is there a role for government in improving the skills base in this area?

I strongly believe that the costs of making these data available (collection, coding, archiving, storage) and in using them (linkage, analysis, interpretation, evaluation and translation) is the responsibility of good government (both State/Territory and Federal). I also feel that this is a vital investment for improving services and the whole activity is likely to make considerable savings because proper use of data should identify wasteful, harmful and costly services which should be terminated, modified, better coordinated etc.

Whilst you can separate out the costs (eg making data available; data management, data analysis and use as you identify above), the ways in which we as researchers obtain funding to obtain, link and analyse data are via government or NGO funded grants. Thus the majority of funding for these activities is from government. NGOs such as the National Heart Foundation or the Cancer Councils in each state may have initiated registries and data collections but most are now state or federally funded as they become statutory collections. This ensures not only their financial sustainability but that they are complete population collections and hence most valuable as described earlier. NGOs and private philanthropic trusts and foundations also fund research grants and sometimes these are collaborative with government funding agencies. One could envisage public private partnerships to fund some of this important work in the future.

We are charged as researchers to access linked data sets but these costs are funded by our grants which we get most commonly from government (NHMRC, ARC etc), NGOs (eg Red Nose Day for SIDS, Rotary Health Fund etc) and sometimes from direct government tenders.

The final questions in this set raise the biggest challenge facing the Federal Government in its desire to use its data more for the myriad of purposes listed in the PM&C Report Public Sector Data Management (2015). There is a huge need for capacity building here; a large, competent, skilled work force in data management, IT, epidemiology, statistics and other analytical capacity is needed. And most of these would benefit from having had exposure to, or advanced training in, the areas of interest (health, education, economics, environment, etc) or teaming up very closely with the bureaucrats in government departments who do. Even if most of the analyses were done by external groups of experts, the upskilling of government public servants is key to them realising the benefits of their data sets and appreciating and understanding the findings, the limitations, the complexity of the interpretations etc. Such upskilling is going to have to be a much bigger investment than that intimated in the PM&C report which suggests that some fellowships for 6-12 month training would be available. To have such training may be fine, but for the fellows to then go

back in to environments which are not supportive and lack capacity in this area means that their skills would quickly wither. Hence my earlier suggestions for several options – outsourcing to groups of competent researchers all around Australia in institutes, centres and universities including NGOs; collaborations between researchers and bureaucrats; government tenders; a national agency like AIHW expanding to take on the role of a Public Statistical Agency etc

QUESTIONS ON PRIVACY PROTECTION

What types of data and data applications (public sector and private sector) pose the greatest concerns for privacy protection?

Data which could embarrass a person or influence their ability to be employed; really any identification of an individual and their profiles or characteristics is an invasion of their privacy. The reasons for doing studies involving such data sets need to be important questions of great good to society.

How can individuals' and businesses' confidence and trust in the way data is used be maintained and enhanced?

What weight should be given to privacy protection relative to the benefits of greater data availability and use, particularly given the rate of change in the capabilities of technology?

Are further changes to the privacy-related policy framework needed? What are these specific changes and how would they improve outcomes? Have such approaches been tried in other jurisdictions?

How could coordination across the different jurisdictions in regard to privacy protection and legislation be improved?

There is a lot of debate about this in Australia; Professor Megan Richardson in the Law School at the University of Melbourne has some excellent research and thinking on these issues.

How effective are existing approaches to confidentialisation and data security in facilitating data sharing while protecting privacy?

See comments below re WA Data linkage.

What lessons from overseas jurisdictions can Australia learn from regarding the use of individuals' and businesses' data, particularly in regard to protecting privacy and commercially sensitive or commercially valuable information?

What are the benefits and costs of allowing an individual to request deletion of personal information about themselves? In what circumstances and for what types of information should this apply?

I feel that this is a really difficult area for the huge data sets that we work on, and would be a logistical nightmare. People die, move and would not respond. It would be prohibitively expensive and for little gain in terms of privacy protection or accuracy of public data sets.

What competing interests (such as the public interest) or practical requirements would indicate that the ability to request deletion should not apply?

The WA experience of using population data on health and more recently other outcomes such as disability, child protection and education, now goes back nearly 40 years. We believe that in spite of having to use personal identifiers to analyse and link data, the methods developed in WA have minimised the risks to privacy being breached. This privacy protecting approach is done within secure servers within the Health Department and the identifying information (name, gender, data of birth, address etc) is only seen by those who set up the master linkage keys. They do not see the sensitive (health, mental health, child protection etc) information but only link to the data sets. The researchers only get the de-identified data linked together. Linkage would be facilitated if Australia

had a system, as in Scandinavia, of unique ID number for every individual in the country. This could actually be done very efficiently by using the Medicare Number in Australia; having such unique identifiers actually protects privacy and enhances confidentiality and results in more accurate linkage.

Page 27

QUESTIONS ON OTHER RESTRICTIONS

I have no comments here.

Page 28

QUESTIONS ON DATA SECURITY

Are security measures for public sector data too prescriptive? Do they need to be more flexible to adapt to changing circumstances and technologies?

Yes and yes.

How do data security measures interact with the Privacy Act?

How should the risks and consequences of public sector and private sector data breaches be assessed and managed? Is data breach notification an appropriate and sufficient response?

Our Telethon Kids Institute has a privacy policy which is very harsh on those who would attempt to breach these data. It is available on our website <http://telethonkids.org.au/privacy-policy/>. It is important that all those who will be using this precious and potentially harmful data understand how it must never be individually accessed or disclosed.

Prof Fiona Stanley AC FAA FASSA

Patron, Telethon Kids Institute

Distinguished Research Professor, UWA

Vice Chancellor's Fellow, UniMelb

June 2016