



National Office: 26 Balmain Cr, Acton, ACT
GPO Box 1956, Canberra, ACT, 2601
(02) 6249 1788
info@socialsciences.org.au
www.socialsciences.org.au

06 November 2020

Ms Deborah Anton
Interim National Data Commissioner
Office of the National Data Commissioner
PO Box 6500
Canberra ACT 2600

Dear Ms Anton,

Re: *Data Availability and Transparency Bill 2020 exposure draft and explanatory materials, and the Accreditation Framework Discussion Paper.*

Please find enclosed the Academy of the Social Sciences in Australia (the Academy) submission to the Office of the National Data Commissioner's *Data Availability and Transparency Bill 2020 exposure draft* and explanatory materials, and the *Accreditation Framework Discussion Paper*.

In 2015 the Academy together with the Australian Urban Research Infrastructure Network (AURIN) hosted two workshops on increasing research access to public sector data and developing a model for trusted access. Please find the workshop reports enclosed which we anticipate will contain further useful information for this reform process.

We commend the Office of the National Data Commissioner for their continued genuine and timely consultation on these reforms which, in our experience, represents a leading practice approach.

The Academy and its Fellows would welcome the opportunity to discuss any of the matters raised in this submission. Please contact Andi Horsburgh, Policy Manager on 0466 123 178, or andrea.horsburgh@socialsciences.org.au

Yours sincerely,

Professor Jane Hall FASSA FAHMS
President



NOVEMBER 2020

DATA AVAILABILITY AND
TRANSPARENCY BILL 2020
EXPOSURE DRAFT

ACCREDITATION FRAMEWORK
DISCUSSION PAPER

ACADEMY OF THE SOCIAL SCIENCES
IN AUSTRALIA SUBMISSION TO:

OFFICE OF THE NATIONAL DATA
COMMISSIONER

Submission to the Data Availability and Transparency Bill 2020 exposure draft and Accreditation Framework Discussion Paper

06 November 2020

The Academy of the Social Sciences in Australia (the Academy) welcomes the opportunity to provide a submission to the Office of the National Data Commissioner's *Data Availability and Transparency Bill 2020 exposure draft* (the Bill) and explanatory materials, and the *Accreditation Framework Discussion Paper* (the paper).

The Academy is an independent, not-for-profit organisation that draws on the expertise of over 680 Fellows to provide practical, evidence-based advice to governments, businesses, and the community on important social policy issues. Academy Fellows are significant users of data for statistical and research purposes and will be users of data that becomes available because of these arrangements.

The Academy welcomes the objective of the Bill to *improve how Australia shares public sector data to drive service delivery, evidence-based policy, research and innovation*.¹ The data sharing scheme, in our view, broadly strikes the right balance between sharing data for the public good and privacy and security protections. If implemented, the reform proposals will make a significant improvement to the quality of research in Australia.

The Academy has been actively engaged throughout the development of the data sharing scheme and made previous submission to the *Data Sharing and Release: Legislative Reforms Discussion Paper (2019)*. This submission builds upon our previous feedback and outlines the Academy's comments on the scheme, with a particular focus on the Accreditation Framework. In making these comments, we are keeping in mind the large numbers of university researchers that might be involved, so it is important that the arrangements are workable.

The Academy and its Fellows would welcome the opportunity to discuss any of the matters raised in this submission. Please contact Andi Horsburgh, Policy Manager on 0466 123 178, or andrea.horsburgh@socialsciences.org.au .

¹ Explanatory Memorandum, Data Availability and Transparency Bill 2020 (Cth).

Comments on the *Data Availability and Transparency Bill exposure draft (the Bill)* and explanatory materials

Key points:

- The Academy encourages further consideration of data which will be excluded under the Regulations. Access to some Electoral Roll and My Health data would hold significant value for research purposes. It is recognised that additional conditions may need to be applied to releases of this data and access may be selective.
- The Academy suggests further consideration of sovereignty issues with respect to data on Aboriginal and Torres Strait Islander people. The Academy will consult with experts and provide further recommendations to the Department on this issue at a later time.

Data Exclusions

The Academy encourages further consideration of data which will be excluded under the Regulations. Access to some Electoral Roll and My Health data will hold significant value for research purposes. It is recognised that additional conditions may need to be applied to releases of this data and access may be selective. For example, an Academy Fellow has been involved in the review of the poor performance of the election polls in the 2019 election. The final report will be released in November and shows that unrepresentative samples is the main reason. Access to electoral rolls to derive sampling frames would help overcome this problem. It would be in the public interest as arguably changes in Prime Ministers and resignations of several Ministers have been based on polls that may have been faulty.

Indigenous Data Sovereignty and Accreditation

As described in the Bill Consultation Paper, greater sharing of data about Aboriginal and Torres Strait Islander peoples raises specific issues regarding data sovereignty and data governance. The Academy welcomes consideration in the Consultation Paper of potential safeguards and advisory mechanisms for data custodians to seek appropriate advice from relevant government and non-government entities, such as AIATSIS. There could also be an accredited data service provider who specialises in this area.

The Academy suggests more work is needed on this aspect of the Bill and we will make a supplementary submission to the Office of the National Data Commissioner after we have had time to engage Aboriginal and Torres Strait Islander Fellows and other stakeholders with particular expertise in this area.

Comments on the *Accreditation Framework Discussion Paper*

Key points:

- The Academy supports the single point accreditation approach for universities, noting some functions can be delegated.
- The Academy welcomes the preservation of existing pathways and mechanisms for data sharing: accreditation arrangements should not override any existing arrangements that work well.
- The Academy holds residual concerns about data custodians' preparedness to release data.

User Accreditation

The Academy supports a single point accreditation for universities even though they are very large organisations, noting that a Vice-Chancellor can delegate to a Responsible Officer e.g. Deputy Vice-Chancellor (Research). We note that the Responsible Officer can further delegate some functions. This seems like a workable arrangement.

It is important to note that university structures are dynamic and that individuals and roles are likely to change over time. Therefore, it would make sense to have a formal review from time to time to ensure the sub-units are up to date.

The sub-units will also meet the accreditation criteria in different ways. It is unlikely there will be a common approach across the University. Consequently, it may be necessary to examine whether individual sub-units meet the accreditation criteria or not. It may well be that some sub-units meet the criteria whereas others need further work. Those satisfying the criteria should not be disadvantaged by those who need further work to meet the criterion.

Internationally, it has been shown that the greatest risk from these types of arrangements is endorsed researchers sharing data with other persons who are not endorsed. They see (and mean) no harm in doing so and don't understand the importance of not doing so. This needs to be emphasised in endorsement procedures.

There may be an argument for having separate arrangements for semi-autonomous or autonomous Research Institutes (i.e. separate legal arrangements, own Board). This could be achieved by allowing more than one responsible officer for organisations in certain circumstances or treating these Institutes as a separate organisation.

Please note that we have not consulted with the university sector directly although some of the Fellows consulted currently hold, or previously held, senior positions in universities.

Accreditation of Data Service Providers

The Academy notes the Bill establishes an alternate pathway for the sharing of government data and *all existing pathways and mechanisms for data sharing will continue to operate unaffected as the*

Bill does not replace or change these arrangements.² We welcome this approach as the proposed arrangements should not over-ride any existing arrangements that are working well, for example, the micro data release arrangements for ABS and AIHW. The Academy suggest the initial focus of release requirements be on 'safe data' where the arrangements can be much simpler. It may take more time to resolve the arrangements where there is reliance on the other 'safes'.

Data custodians: culture change and capacity building

As noted in our 2019 submission:

"culture change will be one of the main challenges and possibly the biggest challenge. The change will not happen overnight. Achieving it will require continual work by the Commission. Staff in the relevant agencies do not regard it as an important part of their responsibilities to share data and are 'hard-wired' to protect their Ministers. Many feel they may be exposing their Ministers through sharing data. It needs to be made clear that it is an important part of their job. In addition to the steps proposed in the Paper, good work on data release needs to be recognised in order to reinforce its importance. "

We have residual concerns about data custodians' preparedness to release data. The Productivity Commission noted "a culture of risk aversion". It appears that the Office of the Data Commissioner does not want to push the data custodians too hard at least initially. There need to be incentives. One way to achieve this could be to produce an Annual Report that lists the data custodians and the data releases that they have supported.

Response to questions posed in the Accreditation Framework Discussion Paper

1. *What is considered to be an appropriate level of Australian ownership for an organisation to be eligible for accreditation?*

Given the sensitivity of individual data and potential security issues, 100% Australian ownership might be reasonable. However, in the university sector, there will be joint projects with other countries which are with consistent the national interest. For international researchers, the appropriate place for endorsement/access decisions lie with the data custodians who will best understand the sensitivities. That is, endorsement may only be for particular data sets where the risks from international access are considered low. For joint projects, it may be the Australian researchers only who can access data about individuals.

2. *Should individuals acting on behalf of an Accredited Data Service Provider be accredited individually? If so, what might be appropriate arrangements?*

This would not be necessary if there are individual confidentiality agreements already in place (e.g. ABS) but may be prudent in other cases.

² Explanatory Memorandum, Data Availability and Transparency Bill 2020 (Cth).

3. *Are there circumstances when it should be mandatory to use an Accredited Data Service Provider for a data sharing project?*

We thought this should be the case unless the Data Custodian had the necessary data sharing capabilities.

4. *What would those circumstances be?*

See answer to Q3

5. *Are there elements of data capability that should be given more or less weight in the accreditation process, i.e. making elements mandatory or optional?*

All three criteria are essential for accredited data service providers. The first two criteria are the most important for accredited users. It will be difficult for the National Data Commissioner to make judgements on the third criteria for accredited users.

6. *What elements would be most useful to Data Custodians to support their decision-making process when considering sharing and access to data?*

They need to understand the risks and may need to ask more questions to understand the risks. At a workshop conducted by the Academy a few years back, it was agreed that safe data remained the best form of data sharing (e.g. ABS, AIHW micro data releases) without needing to be concerned about the other 'safes'. This should remain the focus. However, there will be situations where more detailed data is required and the other 'safes' come into play.

7. *Should the accreditation process recognise other frameworks, standards or processes that have assessed an element of data capability? If so what standards/processes might be appropriate to recognise?*

Yes, if other arrangements are working well now (e.g. ABS, AIHW micro data releases) there is no need to over-ride these arrangements.

8. *Are there any elements of data capability that should be captured in order to understand an accredited entity's ability to keep data safe?*

The three proposed criteria seem to cover the requirements. A proven track record would be useful and, if that exists, it may simplify the assessment for accreditation.

9. *What is a reasonable period of time to assess an application?*

One month is reasonable for accreditation. However, it should be much shorter for requests to data custodians for data releases – one week would be reasonable.

10. *Are there further ways we can streamline the accreditation process?*

We do not have any suggestions at this time, but it would be reasonable to review after 12 months, say.

11. *Do the timeframes to renew accreditation, every 5 years for Accredited Data Service Providers and every 3 years for Accredited Users, seem reasonable?*

Yes.

12. *Is it appropriate to notify parties to Data Sharing Agreements of an accredited entity's suspension?*

Yes, together with the reason and the current status.

13. *Is there any information that must, or must not, be made publicly available through the registers of accredited entities?*

There probably is but transparency is important so some information should be made available. One thing that should not be made public without permission is the contact details of endorsed researchers.

14. *Is there any information that should be made available to Data Custodians through the registers of accredited entities?*

Yes, they should know the responses to the accreditation criteria at least. This should not be kept from the data custodians.

15. *Is charging a fee for accreditation, such as a renewal fee, reasonable?*

Accredited data service providers are providing a service to others – they are not the users but the agents for implementing government policy. Fees should be applied to data users but only if a service is provided. You would not want the situation where a user pays for accreditation but receives no data service. Is a levy on an actual data service a fairer means of cost recovery? Alternatively, the accreditation process could be regarded as a public good.



RESEARCHER ACCESS – CONCLUSIONS FROM 26 MARCH WORKSHOP

Background

The workshop was hosted by the Academy of Social Sciences (ASSA) and AURIN at Melbourne University on 26 March 2015. It was attended by a mixture of Commonwealth Government and Research Institute data providers, as well as research users from Government and Universities. The list of Attendees is shown at Annex 1. An outline of the Workshop is shown in Annex 2. The focus was on unit record or microdata.

There were four parts to the Workshop.

1. Survey based microdata, including recent and proposed developments.
2. Administrative based and linked microdata, including recent and proposed developments.
3. Emerging Services provided by Research Institutions.
4. Data Provider responses to research concerns.

The Workshop did not provide any great distinction between the issues associated with survey data and administrative data. Consequently, most of the comments below apply to both forms of data. Indeed, one of the areas of increasing interest is the integration of survey and administrative data.

The following summarises the main conclusions reached at the Workshop. There was broad consensus on most points.

Conclusions – Limitations on Researcher Access

Despite the increasing availability of microdata sets, and the variety of means of accessing them, there was universal agreement on the need to increase researcher access but in a way that respects the crucial privacy and confidentiality issues, legislative constraints (but it may be important to make some legislation changes), imperfections in data quality and the costs involved in supporting researcher access.

There are considerable potential benefits to Australia from increased access but only if the main findings of the work of the researchers (based on public data and often using public funding) are to become public. The evidence obtained from this research can be important for policy analysis and other purposes and lead to much improved policy and policy implementation. At present in Australia US data is being used too often by researchers because of the easier access and the findings are of limited value to Australia. Furthermore, improved access avoids researchers trying to conduct their own collections usually of a lower quality and often using government funding. From the data provider's point of view, it increases the value of the data being collected. Direct and indirect users of the data can become advocates of the data collection.

The arrangements for the research work can vary and may be (a) work commissioned by policy agencies, (b) work funded by government grants, (c) work based on other sources of funding, or (d) simply 'curiosity' research.

The Workshop showed strong appreciation of the very positive steps undertaken recently by the Australian Bureau of Statistics (ABS) and the Department of Social Services (DSS) to increase access, including the intention to review how legislation is being implemented, the legislation itself, as well as the services provided using NCRIS funding. Other agencies such as AIHW and AIFS have also made important contributions. Researcher access is much better than it was a decade ago. Nevertheless, there were many residual concerns about the current arrangements and the demand has increased particularly for linked data. The most important issues are listed below and discussed in more detail in the following paragraphs.

- (1) Lengthy lags from the request for access to a decision on approval.
- (2) The cost and complexity of the approval process.
- (3) Access to some important data is still not possible.
- (4) A lack of consistency in practices across agencies.

Lengthy Lags

The lags can be substantial especially if the data is not already available and ethics and custodian approvals are required. The following are illustrative examples given at the Workshop but feedback suggests significant lags are involved with most researcher requests for access.

The first example is from the Productivity Commission who was seeking access to administrative data from the Department of Human Services (DHS) and State public housing agencies for a research project that examined the effect of housing assistance policies on employment outcomes. Even though this topic was non-controversial it took the best part of half a year to obtain the necessary approvals. Even then, the Commission was only able to receive 1% samples from DHS, so ended up using a population dataset extracted from the Research and Evaluation Database (RED) which contains the Centrelink Payments data but is curated by the Department of Employment. This project, undertaken over 2014, highlighted the challenges, but also the gains that can be made from cooperative approaches.

The second example is data sought from the Population Health Research Network (PHRN). There was one data access request that involved four jurisdictions, five ethics committees, nine data collections, eight data collectors and five linkage units. Not surprisingly, it took some time to obtain all the necessary approvals and the research loses its impact as a consequence.

Even requests to a single custodian can take a long time to approve. For example, the ABS agreed there was a need for much improved performance in this respect and is taking steps to do this.

Access to purpose built survey longitudinal data, such as HILDA, is less problematic with well-developed protocols for researchers to follow.

Cost and Complexity of the Approval Process

This is related to the lags issue. It can be seen from the above examples that multiple approvals are often required. This can be expensive to both the researcher and the data providers. The multiple approvals in the PHRN example would involve considerable cost in organising the various submissions. A challenge for researchers is the decision about whether to apply for grants before they have the approvals for accessing the data, as lack of certainty about access to data may affect

the provision of the grant. But going through a long and expensive process to access data, without the grant funding to support this process and the work, is also problematic.

Non-accessibility to some data

Although governments have signed an 'open data' policy which is supported by senior officials in government agencies, there is still a lot of resistance within agencies. This makes the agency approval process more difficult and more complex.

Why are the data sets not made available even though there is an open data policy? Apart from concerns about privacy protection, there are multiple reasons such as competing priorities and lack of skills to do the data management work. Also, political sensitivity should not be underestimated. There is an underlying belief by some staff (often middle level) that they may be placing Ministers at risk if more data is released. It is very different to the culture around the release of macroeconomic statistics where 'open government', efficient policy and efficient markets all dictate the release of a wide range of information around the economy. Some statistical releases will be 'good' news for Ministers, others will be 'bad' news but it is not suggested that only good news be released.

Community Views and Consent

There was discussion about community views on the access to microdata for research purposes. Studies have been taken in Australia and elsewhere. They universally show widespread approval especially if it is to be used for health research. The request of the consent of individuals was discussed but was generally not supported especially if respondent co-operation was protected in other ways. It can be expensive and those that do not provide consent may not be typical of the rest of the population thereby distorting the analysis.

Conclusions – Increasing Access

How do you address these issues which are impeding increased access?

- (a) Work to overcome the barriers that limit the amount of data that data custodians will provide.
- (b) More trust needs to be placed in researchers whilst clearly specifying the conditions of access. There is a mutual obligation on researchers to behave in a way that is expected of them.
- (c) Increase the availability of ready to release data sets including already integrated data sets such as those hosted by PHRN and AURIN.
- (d) Acknowledge that different models are appropriate for different users and different data sets. Try to reach consensus on the models and their applicability.
- (e) Provide a common law legal environment which could act as a default when legislation does not already exist.
- (f) Rethink the role of ethics committees and increased mutual recognition of the findings of the ethics committees.
- (g) Find ways of increasing the amount of data that can be released within individual data sets.
- (h) Continue to research technological and methodological solutions that increase access.

Each of these is discussed in turn in the following paragraphs.

Trust in Researchers

By and large researchers want to do the right thing and would accept and comply with reasonable conditions and constraints. A risk management rather than a risk avoidance approach can be justified. The starting point should be that researchers can be trusted but how do you avoid deliberate (very unlikely) or accidental (more likely) breaches of conditions. Some steps that might be taken are:

- (i) A statement on the respective responsibilities of the researchers and the data providers. One important requirement is for researchers to provide applications for data access that provide data custodians with confidence that the data will not be misused and the research has a net benefit. Templates might be developed to assist this.
- (ii) The development of standard protocols for the release of micro data sets, including the licensing arrangements. Individual releases could be based on these protocols as could undertakings to be signed by the researcher. Ideally, these undertakings should be legally enforceable if there are breaches. Model documents could be prepared.
- (iii) Guidelines on how breaches might be managed. These might vary depending on the seriousness of the breach. For example, legal action should be taken where the breach was deliberate and significant. In other cases, the actions might vary from banning future access to the researcher and their institution and a warning.

There was some discussion of a 'Trusted User Model' where the $\text{Probability}(\text{Disclosure}) = \text{Probability}(\text{Disclosure} | \text{Attack}) * \text{Pr}(\text{Attack})$. For access by public servants, there is the Crimes Act, other legislation and the Code of Conduct to ensure public servants are doing the right thing ie make $P(A)$ small. Under the model $P(D) = P(D | A)P(A)$, $P(D)$ will be small if we make $P(A)$ small; and if this can be made to happen for researchers, custodians don't need to do as much on $P(D | A)$, which is what they have been concentrating on hitherto (eg removing matching risks) often limiting the usefulness of data. For academic researchers, the implementation of the three measures listed above should make $P(A)$ small but it requires the co-operation of the research institutions.

Increased Availability of Ready to Release Data Sets

Lags are frustrating and these will inevitably be longer when the requested data sets have to be created. It is much easier if the data sets are already available that can be accessed by any researcher who meets a predetermined set of criteria. This is already happening to some extent in organisations like the ABS, AIFS and AIHW although it is conceded that administrative processes might be improved. DSS has chosen to amalgamate the four longitudinal studies it manages under the banner of the 'National Centre of Longitudinal Data'. It is proposed that within this entity, custodians of the national longitudinal collections will be able to share a range of data management techniques, including those which have the effect of improving data access to researchers'.

More generally, for administrative data, if departments had curated databases of the data that were well documented, specific datasets would be easy to extract. RED is an example of this approach. Having been extracted, datasets should be hosted for reuse wherever possible. This is particularly important for where data from different sources has been linked, as such linkage involves extra effort to create and check the dataset.

AURIN is another example of ready to use data sets and works well from a researcher access perspective. (Note that AURIN doesn't actually provide access to the microdata itself but enables analysis based on the microdata.) The PHRN is not quite the same. Although the data sets are available, many approvals are required before data can be released and this causes inevitable delays.

Different Models for Different Users

The UK has a framework based on the 'five safes' and the nature of the data sets (discussed below). Access arrangements are determined using this framework. This could be adapted for the Australian situation. It already has been by New Zealand using Statistics New Zealand as the co-ordinating authority for the Integrated Data Infrastructure.

Common Law

Some organisations like the ABS and AIHW have their own legislation, although it is being reviewed in the case of the ABS. Most other organisations do not have relevant legislation. There would be benefits in having a common law approach based on best practice. This could underpin the release practices for these organisations. It should also enable greater consistency of practice across organisations.

Ethics Committees

Ethics Committees play an important role but the requirements are onerous particularly for projects that involve multiple jurisdictions. This increases costs and can cause significant delays in the approval process. There were two main concerns. First, ethics committees were sometimes used when ethical clearance may not be necessary given the nature of the request. Second, several ethics committee approvals were sometimes necessary on the same request. Surely, there can be some rationalisation. If an appropriate ethics committee is established, couldn't all jurisdictions act on the advice provided by that Ethics Committee?

Increasing Available Data within a data set

Researchers want to have as many data items as possible available especially if their analysis is exploratory in nature (ie curiosity research). Custodians want to limit the data available and ask researchers to be quite specific about their requirements. One solution may be to provide a sample of the full data set to enable sufficient analysis to be undertaken to allow researchers to be more specific about their requirements. This would also circumvent the significant resources and costs on behalf of researchers, ethics committees and data custodians associated with the provision of additional data that was not specified in the initial approvals.

Technological and Methodological Solutions

These are important and are one of the reasons for increased access over the last decade. There may be further opportunities especially around the linkage of data sets. The ABS is well placed to provide leadership here through its technical and methodological strengths and networks with those doing similar work in other countries.

Linked Data

The Workshop supported the establishment of the Integrating Authorities, namely the ABS, AIHW and AIFS. They are still relatively new so their procedures would not yet be mature. A recent review has recently been completed. It showed that the focus to date had been on protecting the integrity of the data rather than helping researchers to integrate data. Whilst the former is important, there is a need to better balance these two streams of work.

Conceptual Framework for Addressing Researcher Access

Several countries, including the UK and New Zealand, used a conceptual framework based on the 'five safes'. These are:

- Safe people – researchers can be trusted to use the data appropriately and follow procedures
- Safe projects – the project has a statistical purpose and is in the public interest
- Safe settings – security arrangements prevent unauthorised access to the data
- Safe data – the data itself inherently limits the risks of disclosure
- Safe output – the statistical results produced do not contain any results that disclose details about individuals.

This framework could be adopted by Australia. In its presentation, the ABS said it was only currently looking at the first, fourth and fifth 'safes'. Perhaps others should be looked at as well when making decisions on researcher access.

The UK goes further and considers the nature of the data sets when making decisions. For example, is it a sensitive or non-sensitive data set?

Where to from here?

The desired future situation was that there would be improved researcher access to integrated data sets in a way that respects the confidentiality and privacy of the subjects of the data. Unless otherwise agreed, the outputs of this research should be in the public domain for the benefit of those involved in policy development and monitoring, planning of services, and so forth.

It was agreed the main focus should be at the Commonwealth level. Their data sets were more extensive and generally were of greater interest. Procedures at the State level could be adapted from the Commonwealth arrangements. Therefore, it is important that the States have an opportunity to influence the Commonwealth arrangements.

An exception is the health sector where the States have already established good processes. Regardless, harmonisation is desirable with jurisdictions learning from each other's experiences.

The proposed changes will require some modifications to the current arrangements of data custodians even though there have been important steps in the right direction. Importantly, it places an obligation on researchers to behave in accordance with any agreements reached with data providers. If there are breaches there will be consequences for them.

The Workshop agreed that it will not happen without leadership. Where will that leadership come from? The Workshop did not discuss this specifically but the ABS is well placed. It has the technical knowhow and legislated responsibility for the National Statistical Service which covers the States as well as the Commonwealth. In the UK and New Zealand, the leadership has come from their National Statistical Services.

The DSS also has an important role as custodian of the many administrative based data sets and longitudinal surveys. They should also be part of the leadership especially given the changes in policies they have recently been trying to implement. A starting point might be to establish a Working Group to address the issues associated with increased access. The ABS and DSS would be key players.

Infrastructure funding through NCRIS has been important but it is suggested future funding priorities should be based on research projects using integrated data sets.

In summary, the keywords for improvement are trust, leadership and culture.

MAY 2015

ANNEX 1

Details of workshop

- Objectives, agenda, etc
- List of participants



ASSA Roundtable – Trusted Access Model

Prepared by Dennis Trewin

Held on 9 November 2015.

A list of Attendees is at Attachment 1 and the Agenda for the day is at Attachment 2.

Background

There is a strong and growing interest across government in maximising the use of public sector data for policy and research. While governments seek researcher expertise to analyse complex social and economic issues, at present access to public sector data can be hampered by barriers on both sides.

In March 2015, ASSA together with the Australian Urban Research Infrastructure Network (AURIN) conducted a workshop on increasing researcher access. The workshop recognised that by and large, researchers want to do the right thing and would accept and comply with reasonable conditions and constraints. Further it argued that a risk management rather than a risk avoidance approach could be justified. The starting point for developing proposals should be that most researchers can be trusted but how do you avoid deliberate (very unlikely) or accidental (more likely) breaches of conditions?

The March workshop recognised that some steps that might be taken toward safe arrangements for microdata access are:

- (i) A statement on the respective responsibilities of the researchers and the data providers. One important requirement is for researchers to provide applications for data access that provide data custodians with confidence that the data will not be misused and the research has a net benefit. Templates might be developed to assist this.
- (ii) The development of standard protocols for the release of micro data sets, including the licensing arrangements and certification process. Individual releases could be based on these protocols as could undertakings to be signed by the researcher. Model documents could be prepared.
- (iii) Guidelines on how breaches might be managed. These might vary depending on the seriousness of the breach. For example, legal action should be taken where the breach was deliberate and significant. In other cases, the actions might vary from a warning to banning future access to the researcher and their institution.

The second workshop held in November 2015 built on the foundation laid by the first. Its purpose was to continue the discussion on how to improve researcher access to public sector data under what is referred to as a Trusted Access Model. This workshop also broadened the scope of the discussion by including participants from a larger number of Commonwealth government agencies. A keynote speaker, Felix Ritchie, from the UK and



formerly the UK Office of National Statistics was invited to address the workshop about international developments in data access.

Trusted access and the five safes framework

The intention of trusted access is to safeguard privacy and confidentiality through well-structured partnerships rather than heavy data confidentialisation or severe restrictions on access. Trusted access could be based on a framework known as the '5 safes' which has already been adopted in the UK, parts of Europe and New Zealand. The basic premise of the framework is that data access can be seen as a set of five risk or access dimensions:

- safe people
- safe projects
- safe settings
- safe data
- safe output

The key to the framework is that the five dimensions independently and in combination contribute to consideration of whether a particular instance of data access meets expectations for privacy and confidentiality. Steve McEachern from the ANU observed in his presentation that the framework was like a graphic equaliser with a slider for each dimension.

Trusted access to micro or unit record data, therefore, may be implemented as one in which the people, purpose, settings and output dimensions are heightened and in which the data dimension is reduced. This may be contrasted with general access to, say, aggregate data where there is more complete confidentialisation of the data and the other dimensions do not need to be addressed at all.

Settings for trusted access

Safe People

Can the researchers be trusted to use the data in an appropriate manner?

The workshop recognised the shift in government policy to more open access and to a more risk management approach to the provision of unit record data. Participants acknowledged that this is an opportune time to build partnerships between government and the research community that acknowledge the mutual benefit of researcher access to data and an opportune time to develop mechanisms for shared accountability.

A key consideration for the safe people dimension discussed by participants was that data custodians set clear expectations and researchers understand and practice their responsibilities. Participation in training of information sessions could, for instance, be mandated before a researcher can be regarded as 'safe'. Legislation may also be required to support undertakings to be signed by researchers and to deal with breaches. Whilst recognising that there may be some legislative limitations, participants supported efforts



directed toward streamlining the authorisation of researchers to access data from government agencies.

Correspondingly, government agencies could adopt risk mitigation practices to ensure researchers are safe people. These might include:

- establishing the bona fides of the researcher, taking account of the researchers previous history in accessing data if available;
- provision of a training module (which may be on-line);
- entering into a legally binding agreement that sets out responsibilities for both partners; and
- an emphasis on communication rather than policing.

Safe Projects

Is the data to be used for an appropriate purpose?

In the context of the five safes framework, the main focus of this dimension is whether the data are being accessed for statistical rather than compliance or, perhaps, commercial purposes. As it is public sector data that is being accessed, the question arises about the extent to which anticipated public value should be assessed when assessing an access request from a researcher.

This dimension of the framework generated significant discussion at the workshop. Some researchers thought that a requirement to provide information about the purpose of the project could result in the independence of research being compromised, for instance, if government agencies deny researchers access if they thought the results could reflect unfavourably on the agency or the Government. Participants all agreed that this situation had to be avoided.

Researchers identified a number of professional controls on research including the Australian Code for the Responsible Conduct of Research (ACRCR), the human Research Ethics Committee (HREC) and other professional and journal requirements, all of which could be seen as moderating the need for data custodians to have detailed information about the project.

It was noted that, where relevant, the UK use the Administrative Data Research Council to approve safe projects and Germany have a similar vetting authority The Institute of Employment Research (IAB).

It was suggested that an important related question was whether or not the data was fit for the purpose for which it was being requested.

Safe Settings

Does the access environment prevent unauthorised use?



The workshop agreed that the goal should be systems that deliver microdata access to the desktop in preference to secure locations such as onsite data laboratories.

Safe Data

Is there a disclosure risk in the data itself?

Except in rare circumstances, participants agreed that unit record or micro data would have a first level of disclosure risk managed through removal of personal or business identifiers such as name and address. However, further confidentialisation such as through aggregation or masking introduces a trade-off between disclosure protection and analytical utility. Participants noted that once there is established trust and a structured process for authorising access, the risk of data being attacked may well be reduced and privacy and confidentiality could be protected with a lower amount of direct confidentialisation.

Participants noted that there are mutual benefits here to government (reduced time and resources) and to researchers (improved timeliness and utility of data).

Safe Output

Are the statistical results non-disclosive?

Vetting of output is a safeguard to ensure that confidentiality is maintained in the results of the research. There was some concern that output checking may slow down the research or add cost. These concerns deserve consideration. They have been addressed in the UK through high quality training programs that not only teach researchers how to confidentialise output but result in efficiencies in checking, since the material submitted is generally well prepared. Another potential option is to explore automation of checking processes, and there is currently research in progress including in Australia, that may help.

Discussion points

The workshop enabled discussion on a range of issues in addition to those raised against each dimension of the five safes framework.

As noted above, the five safes should be considered in combination and on a sliding scale. They will vary according to the type of access and data. For example, some work may be required to develop arrangements for access to linked unit record datasets which include data from different custodians.

The 'safes' should be interpreted in an Australian context. Trusted access could be refined using experiences of New Zealand and the United Kingdom to inform a system suited to the Australian situation.



The workshop emphasised that the current emphasis on open data was a window of opportunity that should be exploited.

It was recognised that any multi-departmental exercise to implement a trusted access model for data access, especially linked microdata, to researchers must have clear oversight to monitor the process and to evaluate its implementation, efficiency and effectiveness. Ideally, this should be based on existing governance arrangements. Regular audits of the effectiveness of the arrangements should also take place. This would help to provide some public assurance of the integrity of the arrangements.

The workshop recognised the great strides the Australian Government and many of its agencies are making to facilitate better data access. It would be desirable that the public be informed of these developments and engaged for comment, especially with an emphasis on the benefits.

It is important that there is transparency in arrangements for researcher access to the information that the public has supplied through the various administrative processes of government. It would be beneficial to gain support from the privacy and information commissioners and Australia's Chief Scientist.

Trusted access trials

There was consensus at the workshop to progress a trusted access model for researcher access to public sector data based on the five safes framework. Participants agreed that the time is right to undertake limited trials of such a model with the objective of using it more extensively.

Two broad trials were proposed:

- Access by researchers engaged by the Department of Social Services to linked datasets containing welfare payments information and Census Data.
- Access by researchers engaged by the Department of Industry, Innovation and Science to the Expanded Analytical Business Longitudinal Database containing taxation and survey data for Australian firms.

The next step in this process is for collaboration among the participating agencies and researchers to scope and initiate the trials. Remote access was considered a high priority especially the use of virtual data laboratories. Scalability is a very important consideration once agencies go past the trial stage.