ACADEMY OF
THE SOCIAL SCIENCES
IN AUSTRALIA

TRUSTED ACCESS MODEL

ACADEMY PAPER

Trusted Access Model

Dennis Trewin FASSA

Academy Papers 4/2016
Roundtable Report

The Academy of the Social Sciences in Australia
Canberra 2016

Academy Papers
ISSN: 2202-4697 (Web)

Requests and enquiries concerning reproduction rights should be addressed to:
The Academy of the Social Sciences in Australia
GPO Box 1956, Canberra ACT 2601
T +61 2 6249 1788 • F +61 2 6247 4335
Email: secretariat@assa.edu.au.

Contents

Trusted Access Model

This Academy Paper is the outcome of two policy Roundtables convened by the Academy of the Social Sciences in Australia in conjunction with the Australian Urban Research Infrastructure Network (March 2015) and the Australian Bureau of Statistics and the Department of Social Services (November 2015).

Roundtable Convener

Dennis Trewin FASSA

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 access 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.