Productivity Commission Inquiry into Data Availability and Use: Submission Australian Urban Research Infrastructure Network (AURIN)

AURIN commends the Productivity Commission for its interim report and we appreciate the opportunity to present at the Public Hearing on 21 November in Melbourne. The interim report identifies the growing importance of data release in the Australian economy and the consequential need for infrastructure and expertise to facilitate management of this. AURIN has delivered a proven, innovative platform that links researchers and government to data, infrastructure and analytical tools to support evidence based decision making. As such we are positioned to facilitate the role of data in lifting productivity.

- AURIN https://aurin.org.au/ provides a complete data release and delivery service for researchers and government (local, state and federal). AURIN coordinates the collection of data from a variety of authoritative data sources (public and private sector), and provides access to data through the AURIN Portal and associated sites like AURIN Map http://map.aurin.org.au/

- AURIN's experience providing a complete data release and delivery service would position it well to operate as an Accredited Release Authority (Draft Recommendation 9.6). Since its inception in 2010 AURIN has coordinated the release of around 1500 datasets for access and use by the research and government community and has strong relationships with data custodians in industry and government. AURIN consults regularly with well its developed national researcher networks and end user groups to identify datasets of national interest for research and decision making.

1. AURIN operates as an Accredited Release Authority (ARA)

The interim report proposes the selection of state/territory agencies for accreditation as Release Authorities with oversight from the proposed National Data Custodian (Draft Recommendation 9.5). AURIN's experience providing a complete data release and delivery service would position it well to be an Accredited Release Authority.

A key criteria for selection as an ARA is "sectorial expertise" (Draft Recommendation 9.6)*.* AURIN could operate as an ARA for the tertiary sector due to its well developed researcher network and experience managing the release of datasets through the AURIN infrastructure since 2010.

Under Draft Recommendation 9.6, Accredited Release Authorities would be responsible for:

- deciding whether a data set is available for public release or limited sharing with **Trusted users**; and
- Curating and ensuring updating of **National Interest Datasets.**

National Interest Datasets (NID's) are datasets that are designated (following consultation with state and territory governments) to be of national interest and importance for public access or, if sensitive, restricted access by trusted user. (Draft Recommendation 9.4).

Trusted users would be accredited by a new body National Data Custodian for access to restricted NID's and could include a range of entities including universities and research bodies covered by privacy legislation. (Draft Recommendation 9.7).

Current AURIN activities that align with Draft Recommendations include:

- Negotiating access to public and private sector datasets of national significance eg: national sales evidence in response to demand from its end user researcher community to carry out Australia wide research; and
- Arranging for limited release of government and private sector datasets restricted by legislation or contractual arrangements to approved end user groups.

2. ***AURIN provides a complete data release and delivery service for researchers and government***

AURIN negotiates access to high quality data and also provides a secure purpose built platform, the AURIN Portal, through which to access these data. Policy makers and researchers are able to use the data to gain insights and develop policy on shared and reliable data. The results of the research can also be released back to the AURIN end user community through the AURIN Portal and other AURIN platforms such as AURINMap http://map.aurin.org.au/ AURIN has extensive experience in:

- **Managing the release of high value data** including managing access to restricted or sensitive datasets. Examples include the Australian Business register data (geocoded business addresses) and the National Health Services directory data (locations of GP's) and property data from Corelogic and Australia Property Monitors currently being negotiated;
- **Developing metadata** including current collaboration with the Australian National Data Service (ANDS) to develop a metadata tool and expand metadata available about datasets via Research Data Australia (RDA);
- **Building relationships with data custodians** in government and the private sector https://aurin.org.au/resources/becoming-an-aurin-data-provider/ ;
- **Developing and advising on data standards** including approaches to metadata and licensing.

3. ***AURIN comments on other Draft Recommendations/ Information Requests***

We offer two additional comments on pages 36 and 38 of the draft report:

Information Request – page 36

"The Commission seeks further views on datasets that are of national interest and that could feasibly be designated as such under the proposed process"

Australian researchers urgently need data which is spatially enabled. This means that the data can be routinely accessed and able to be mapped. Over the years AURIN researchers have identified themes of datasets which could be released these themes include land use (property values and sales evidence), economic activity, education, health, social, justice, environment, economic and infrastructure (3D buildings and land models). Stakeholders have identified that it is most beneficial to research when it is available in as detailed form as possible (for example, unit record level).

A number of these datasets are available are released in ad hoc ways or are available on web-sites.

AURIN also supports:

- Facilitating the spatial linking of data by releasing aggregated data according with the ABS ASGC standards. Currently many data collections are via postcodes and the incompatibility of the systems causes difficulties in the analysis of the data.

- Utilising existing technology to improve the quality and currency of data (and associated metadata) at a fine scale (resolution), so the data is able to be applied for research, and importantly, leverage technologies that overcome data heterogeneity.

Information Request – page 38

 "The Commission seeks further views on the establishment of a Parliamentary Committee to take community input on possible National Interest Datasets, to review nominations made, and make proposals for future designations. View are also sought on practical alternatives."

We support this approach and welcome the opportunity to be involved.

Draft Recommendation 6.1: AURIN supports development of uniform data standards to make data "fit for release" including license information

Draft Recommendation 6.1 lists a series of objectives for the standards. We think that data licensing information/approach should be included as part of data standards. AURIN experience is that obtaining consent or a licence to use and share data is an important element in assessing whether data is fit for release.

This is particularly relevant in the research context where many projects producing data are collaborative. Questions around who or what institution will ultimately control/release the data, whether it can be published and by who, who is attributed as the creator etc. are often raised as are questions around data format.

Obtaining consent of creators to licence or release data may still be relevant even under open data licence regimes. Being able to identify the person who created the data or the copyright holder is required for most Creative Commons licences to operate.

Open licences do also contain include some minimum obligations for data users to comply with eg: obligation to attribute the creator. We think that it makes sense to include licensing information and approach as part of a coordinated data management strategy that makes users aware of data standards including any licence conditions.

Commonwealth Funding agreements often refer to/or mandate open data release or open source release and data sharing but don't provide a clear framework for this. It would be useful if a practical coordinated approach to data standards and open data licence release could be recommended upfront. In this way participants could actively agree to share data outputs of research before research begins.

Ideally research data could be shared through an open data licence or if sensitive released to a specific group to further research. AURIN experience is that these issues need to be addressed in parallel with data standards around access and usability.