

29th July 2017
Data Availability and Use
Productivity Commission
GPO Box 1428
Canberra City ACT 2601

Re: Place, Health, and Liveability Research Program submission in response to
the Productivity Commission Inquiry into Data Availability and Use.

To whom it may concern,

Thank you for the opportunity to make a submission to the Productivity
Commission Inquiry into Data Availability and Use.

Yours sincerely,

Prof Billie Giles-Corti
Director, McCaughey VicHealth Community Wellbeing Unit and on behalf of the
National Liveability Study and Place Health and Liveability teams

Productivity Commission Inquiry into Data Availability and Use

Submission from the National Liveability Study and Place, Health and Liveability research groups

Who we are

The Place, Health and Liveability Research Program, located in the McCaughey VicHealth Community Wellbeing Unit, Melbourne School of Population and Global Health at the University of Melbourne, studies neighbourhoods to understand the relationships between built and socio-cultural environments and health. Healthy and liveable communities provide the basis for social equity, harmony, economic resilience and environmental and social sustainability. Our research aims to inform policies and practices to create healthy, liveable communities and our submission reflects an evidence-based approach to achieving this goal.

The Place, Health and Liveability Research Program leads the NHMRC Centre of Research Excellence (CRE) in Healthy, Liveable Communities and the National Liveability Study, funded by the NHMRC Australian Preventive Partnership Centre. These research programs include collaborative research with universities across Australia and internationally, including the University of Western Australia, Australian Catholic University, University of Queensland and the University of Wollongong, as well as Cambridge, Harvard, the University of British Columbia and University of Calgary.

The Place, Health and Liveability Research Program also works alongside Community Indicators Victoria (CIV), supporting the development and use of local community wellbeing indicators across Victoria to improve citizen engagement, and evidence-based community planning and policy making. CIV makes data from multiple sources available at Local Government Area level for all Victorian municipalities and also provides customised small area data at the neighbourhood level. CIV indicators are used by local, state and federal government departments, community organisations, schools and philanthropic organisations. These data are provided to measure societal progress, encourage evidence based planning, monitor improvements over time, engage communities and governing organisations and recent research confirms that CIV is one of the most frequently cited source of evidence in Victorian Municipal Public Health Plans. Data are made available to all members of public to generate health and wellbeing reports that include data on social, economic, environmental, cultural and democratic issues for local areas.

How we use data

Our research investigates relationships between the environment and health and wellbeing. We use our findings to inform policy and practice relating to the design of neighbourhoods. For instance, for some time, we have been examining how neighbourhood walkability, access to green space, shops and services impacts walking and cardio-metabolic risk factors. We are also conceptualising 'liveability' through a health lens, and creating, and studying the impact of, a health-related liveability index.

A key way we disseminate our findings to inform policy and practice is by developing policy-related indicators of liveability, walkability and community wellbeing, which allow for us to benchmark and

monitor the implementation of urban policies. These indicators are made available to other researchers, policy makers and practitioners, and the general public via the Community Indicators Victoria website (<http://www.communityindicators.net.au>). There are plans to make national indicators of liveability – from a health perspective - available for major cities over the next year.

Our work is topical, policy-relevant and generates considerable public interest. Having access to good quality public and private sector data across Australia is therefore essential for our research, knowledge translation and indicator creation. Without appropriate data we are unable to: 1) investigate relationships between the environment and health; 2) provide evidence on these relationships to policy makers, practitioners and government agencies; and 3) create and disseminate indicators of liveability and community wellbeing.

We primarily use two types of data: geographic data that describe the social, built and natural environments (e.g., location of roads, parks, shops), and data on health-related behaviours and outcomes of individuals (e.g., individual health survey data, hospital admissions). We address these two types of data separately.

Importantly, from a productivity perspective, the higher the quality of these data, the less work is required by research teams to prepare the data for analysis. Research on the built environment and health is rapidly growing internationally: an extraordinary amount of time and research dollars are being spent on ‘cleaning’ and compiling datasets because of the lack of high quality national data for research in Australia. This is reducing our research analysis productivity and international competitiveness; and wasting research dollars in a resource-constrained environment.

High value geographic data describing the social, built and natural environment

Characteristics of high value geographic data

From our perspective a high value dataset is consistent nationally, of high quality (i.e., is reasonably accurate), is kept up to date with snapshots at past time periods also available, available from an authoritative source and has metadata. A dataset without these characteristics is of limited use and/or requires significant time and money to make it useable (e.g., manually checking and fixing inaccurate location data). We also consider that high value datasets are those that can be used in a large proportion of our research. For instance, road, land parcel, address and land use data form the basis of many of our measures of the environment. Table 1 lists examples of geographic data – of varying quality - that we have been able to source nationally.

Table 1. Examples of valuable geographic data.

Data	Availability	Cost	License	Historic data readily available?	Source
Population (census)	National	Free/minimal cost	Creative Commons	Yes	ABS
Cadastrre (land parcel)	National	Cost	Varied	No	Local government, State government, PSMA (via private resellers or AURIN)
Address (G-NAF)	National	Free	Adaptation of Creative Commons	No	PSMA
Road centrelines	National	Free/Cost	Varied	No	Local government, state government, PSMA (private resellers), community organisation
Public transport data (GTFS)	Selected states	Free	Varied	No	Government transport agencies.
Health service locations	National	Free/minimal cost	License difficult to obtain	No	National Health Services Directory
Education facility locations	National	Free/minimal cost	License difficult to obtain	No	Australian Children's Education & Care Quality Authority
Features of interest (e.g., location of shops, community facilities)	National	Cost	Restrictive license	No	PSMA, ¹ private companies (e.g., SENSIS, Pitney Bowes)
Land use (for each land parcel)	Selected states	Free/minimal cost	License difficult to obtain	No	Valuer General data on land use for each parcel. No national source.

¹ NB: These data are provided by states, contain inconsistent attributes and require considerable cleaning to unify and create national datasets.

Table 2 lists geographic datasets that would be valuable for our research, however we have not yet been able to source these nationally. It is expected that the upcoming Public Sector Mapping Agencies (PSMA) Geoscape data product will provide some of these data. However, we are unsure at this stage whether the costs will be affordable and the license agreements flexible enough to justify the costs.

Table 2. Examples of potentially valuable geographic data that is not currently available nationally, and in some cases is not available for cities/states.

Dataset
Parks
Footpaths
Building footprints
Tree canopy
Bicycle lanes
Playgrounds
Playgroups
Air pollution
Water quality
Crime locations

Examples of government initiatives that have improved geographic data access and use

Australian Bureau of Statistics

The ABS provides high value, authoritative geographic data on the population. We use this data extensively and are able to do so easily and at minimal extra cost because this data is high quality, easy to download, is in appropriate formats, free/minimal cost with a non-restrictive Creative Commons license agreement and is available historically (i.e. at different time periods). The availability of these datasets have enabled us to use geographic data to improve our understanding of: the liveability of an area [1]; access to transport; prevalence of family violence; access to greenspace, among many others. Of even greater importance is the consistent usage of ABS data in our research for exploring the impact of socio-economic differences on individual health [2, 3] and monitoring equity issues in implementation of policy [4], and creating and disseminating indicators of liveability [5] and community wellbeing [6].

Open G-NAF (national address data)

The recent release of open G-NAF data has improved our ability to use this data in a wider range of health research. Prior to the release of open G-NAF we purchased G-NAF under an educational license. This license meant that we were only able to use the data for a single research project. For each additional project we had to re-purchase the dataset. The newly open G-NAF as a less restrictive Creative Commons license that enables us to geocode participant residential addresses and calculate measures of the environment for a population health surveys without having to find additional funding to purchase address data. For example, we are currently using open G-NAF to geocode residential addresses and calculate measures of the environment for the Ten to Men study – a national survey of Men’s health. This will enable us to investigate how the environment can support or hinder men’s health. If G-NAF had not been made open source, this research may not have been conducted since we would’ve had to have applied for funding to purchase the address data to enable the research.

Open G-NAF also enables us to create our own geographic data describing the environment since address data is used to geocode addresses of a variety of destinations where geographic data does not exist or is difficult to obtain (e.g., library locations).

We will use open G-NAF extensively in the future for geocoding of various geographic and health survey datasets, expanding the range of research we are able to conduct.

Australian Urban Research Infrastructure Network (AURIN)

The Australian Urban Research Infrastructure Network (AURIN) is a federally funded initiative to implement a portal that facilitates access to, and use of, geographic data. Where possible we source our geographic data from AURIN since AURIN provides a number of valuable services, in particular: sourcing datasets (including difficult to access datasets) and negotiation of license agreements. Some of the current limitations of AURIN – provides data as is, tends to provide current data as opposed to historic data – also represent potential opportunities. AURIN could potentially be of even greater value to research if it was able to address these limitations. The already significant investments in AURIN could increase research productivity further by investing in the capacity to clean geographic data and archive historic data.

Issues

Despite these developments, **there are a number of issues that impact research productivity and the cost of undertaking research: the ease of access and licensing data; the cost of data; lack of standardisation of data; the need for data from multiple data sources.** These are now considered:

Data access and licensing

Many datasets are difficult and time consuming to access in an appropriate format. Frequently, substantial time and money is spent negotiating with public sector data custodians, justifying data use, and negotiating license agreements.

Different data custodians and suppliers have different license agreements for the use of their data in our research. This necessitates the time consuming process of renegotiating agreements for each data custodian. Some standardisation of the data licenses would reduce this barrier to accessing and using data.

Some datasets have restrictive license agreements that do not align with the way academic research works: the different time frames (multi-year highly related projects), ways of working (frequently in multi-institute collaborations), focus (less time spent working with the data than a typical business), and requirements (e.g., the need to keep data to support publications that arise from research).

Cost of data

While free data is useful to our research, it is only useful if it is high quality data. We recognise that it is expensive to develop and maintain data and therefore are prepared to pay for data if it is of good quality, requires minimal cleaning, is easy to obtain, and has less restrictive license agreements.

Standardising data

Lack of standardised geographic data can add to the amount of data cleaning we have to do. It also limits our ability to explore the health benefits of liveability nationally. For example, data available

through PSMA is derived from each state and territory, and is collected and attributed differently. We strongly support the standardisation of core, high value, datasets.

Multiple data sources

Many geographic datasets that describe the environment have a number of sources and sometimes we use both public and private sources for datasets that measure the same aspect of the environment. For instance, road centerline data is an essential dataset that we use to model people's potential movement through the environment. We currently source road centerline data from a variety of sources: local governments (e.g., Brisbane City Council), state governments (e.g., Victorian Department of Environment, Land, Water and Planning), private companies (e.g., resellers of PSMA Australia's Transport & Topography dataset or MapInfo's StreetPro dataset), the Australian Urban Research Infrastructure Network (AURIN; who provide access to a number of road datasets), and community organisations (e.g., OpenStreetMap).

Depending on the source, these data can be free or have a cost associated with them. Different sources also have different licensing, ranging from open data to data restricted to specific projects (typically data sourced through private companies).

Benefits of increasing the availability of high value geographic data

Our research would benefit from improving the quality and consistency of *national* geographic data. This would enable us to better use public money and spend less of our research budget on searching for and cleaning data. Less restrictive license agreements, particularly for PSMA and private sector data would also be beneficial if it allowed us to reuse the same data for a range of projects and to share our value added data (e.g., data we have cleaned) with other researchers. Less time and money spent on data-related costs would allow us to spend more time and money on research and developing and disseminating indicators. For the community, more research time means that we will have a better understanding of the relationship between the environment and health and will improve the evidence base from which to inform policy and practice. It also means that we will be better able to develop and disseminate indicators for use by other researchers, policy makers and practitioners, and the public.

There would also be benefits to increasing the range of available geographic data. Despite considerable interest from government and the community in our research, currently we are only able to research those aspects of the environment that we have data for and can measure. The lack of availability of certain datasets means that we cannot provide evidence that policy makers and practitioners need to design healthy, liveable communities. For example, we measure walkability based on the road, rather than the pedestrian, network because we lack data on the location of footpaths and pedestrian trails. This means we are unable to provide evidence on the health benefits of for example footpaths, and are unable to accurately measure the implementation of government efforts to improving features of the built environment which is currently being called for in the *Plan Melbourne Refresh*; and the Federal Government through its Smart Cities agenda.

Recommendations for geographic data

- 1. Open core datasets such as PSMA Transport & Topography and Geoscape. For much of the geographic data that describes the environment, there should be minimal privacy/ethical concerns.**

- 2. Develop national standards for data and create standardised national spatial data sets where appropriate.**
- 3. Where possible standardise licensing of geographic data using Creative Commons licenses, or adaptations of Creative Commons licenses.**
- 4. Archive (and make available) historical geographic data.**
- 5. Consider funding an organisation such as AURIN to manage the cleaning, archiving, and dissemination of geographic data.**

Health data

Numerous data is collated on primary health care including GP visits, specialist consultations, diagnostic tests and pathology. These specific health datasets offer a cutting-edge way forward in understanding how the environment impacts health. The most commonly used sources of health data include the Medical Benefits Scheme (MBS), the Pharmaceutical Benefits Schedule (PBS) and in some instances hospital administration data sets. The MBS and The PBS require consent of the patient, which is limited to a five year window which may not be of use in exploring long run effects of long term health related interventions or exposure to environmental conditions. Furthermore, a limitation of these datasets is that it requires patients to claim on Medicare or to fill out a script. Consequently, these datasets do not offer complete information and do not include information from people who do not claim Medicare.

Hospital data is usually more complete but not necessarily available nationally since not all states have accessible and linked datasets, currently only NSW and WA have linked records.

Recommendations for health data

- 1. All health datasets be made available at a national level.**
- 2. Extend the window of consent to allow for long term research projects.**

Data linkage

The linkage of Australian survey data to medicare data, hospital records and geospatial data would allow research examining the impact of city planning on chronic and respiratory disease; mental health outcomes.

The productivity of research would be enhanced enormously by stream lining the ethics application process to use existing data. National studies are required to apply to multiple ethics committee (often more than one in each state) before undertaking any research. This is excessive, costing time, resources and energy and importantly reduces the funding available for the research itself.

Any attempt by government to improve data sharing, should still be flexible. For example, it is still important that researchers have access to unit records, once they have appropriate ethics approval and safe guards in place. In efforts to make data more freely available, it is important not to 'throw the baby out with the bath water' and ensure there is still a flexible approach that does not preclude important research requiring individual unit record files; in an attempt to meet the needs of most people who do not require these data.

Recommendations for data linkage

1. Establish Linked Data services similar to that offered in Western Australia.
2. Link medicare data with hospital records and geocoded patient outcome data. Where possible obtain participant consent to link the Australian survey data to these data; and facilitate these data being linked to geospatial data.
3. Streamline processes to obtain ethics approval to use data.

References

1. Badland, H., et al., *How Liveable is Melbourne?* Conceptualising and testing urban liveability indicators. Melb.: University of Melbourne, 2015.
2. Sugiyama, T., et al., *Can neighborhood green space mitigate health inequalities? A study of socio-economic status and mental health.* Health & place, 2016. **38**: p. 16-21.
3. Turrell, G., et al., *Can the built environment reduce health inequalities? A study of neighbourhood socioeconomic disadvantage and walking for transport.* Health & place, 2013. **19**: p. 89-98.
4. Mavoja, S., et al., *Area-Level Disparities of Public Open Space: A Geographic Information Systems Analysis in Metropolitan Melbourne.* Urban Policy and Research, 2015. **33**(3): p. 306-323.
5. Giles-Corti, B., et al., *Reconnecting urban planning with health: a protocol for the development and validation of national liveability indicators associated with noncommunicable disease risk behaviours and health outcomes.* Public Health Res Pract, 2014. **25**(1).
6. *Community Indicators Victoria.* Available from: <http://www.communityindicators.net.au/>.