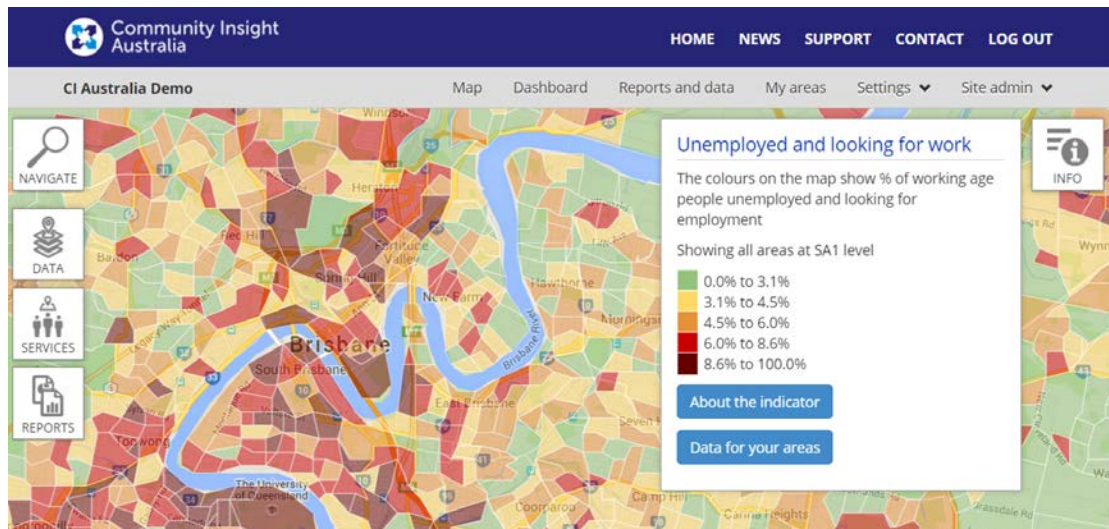


# THE VALUE OF SMALL AREA DATA

## COMMUNITY INSIGHT AUSTRALIA: ENABLING ACCESS TO PUBLIC DATASETS

Community Insight Australia (<https://communityinsightaustralia.org/>) is a new software that enables users to visualize public and private datasets. It allows users quickly and easily to access location-based information for location-based decisions.



In Australia, it is currently difficult to gather data about any given community, let alone several communities at once. Community Insight Australia makes a large collection of data easy to access, understand and use. This collection includes all publicly available data that can be mapped, in combination with an organisation's own datasets. The software also allows users to draw around an area on a map and see all the information available on the people who live there. Information is provided in heat-maps, comparative dashboards and detailed reports, all of which are designed to be used and reused. The software is particularly useful for targeting services to particular communities, and tailoring those services to the needs of people who live there.

	Population aged 0 to 4	Working part time	Employed away from home	Unemployed and looking for work	Weekly household earnings	Weekly household expenses	Highest year of school	People with a Bachelor degree	Attending University	Rented Total	No Internet connection	One parent family	Persons with need for services	Motor vehicles per dwelling
<b>Capital Cities - the most central SA2</b>														
Adelaide City	2.2	28.6	4.6	11.2	36.3	2.8	1.7	23.9	56.0	65.0	12.7	12.0	2.3	38.2
Brisbane	2.4	23.7	4.0	6.7	19.2	6.6	0.7	25.5	36.2	54.1	5.6	8.0	0.9	30.1
Darwin	3.7	15.4	6.2	3.9	16.7	6.5	2.3	16.2	15.6	67.5	12.7	8.5	2.3	15.7
Hobart	3.7	31.5	5.2	5.4	35.5	3.0	2.1	24.0	35.3	52.8	16.8	14.5	4.1	17.8
Melbourne City	2.0	23.4	4.0	10.8	29.2	2.9	0.6	28.7	54.5	66.1	6.6	7.8	0.6	68.5
Perth City	3.2	20.4	4.7	5.2	23.6	6.4	2.1	26.3	26.0	61.3	12.9	7.9	2.4	21.6
Sydney City	2.5	30.2	5.0	8.0	25.2	6.6	1.2	25.2	29.8	63.2	8.6	9.5	1.3	56.3

The platform is designed to serve subscribers who make important decisions about services for communities. It does not require any statistical skills to operate. If they can read a map, they can access all the information we have.

Community Insight Australia is a translation of the English tool, Community Insight, which was launched in 2013. The Australian version is currently in its Beta version and foundation subscribers include government and not-for-profit organisations delivering services to vulnerable people (<https://communityinsightaustralia.org/join-us/foundation-partner/>).

Datasets of high value to our subscribers are:

- **Aggregated by small area:** it is not necessary to know the addresses of individuals to make evidence-based decisions for communities
- **Granular:** the smaller the granularity, the higher the value for decision making
- **Easily accessible:** if datasets have a hyperlink, they can be checked by anyone who wants to verify their origin and accuracy
- **Related to the most vulnerable people:** the more vulnerable the person, the more they benefit from providers better serving them.

Community Insight Australia operates as a social enterprise and from our perspective we have several recommendations that will enhance the availability and use of data in Australia. Our recommendations are based on our own experiences over the last two years.

## RECOMMENDATIONS

---

### 1. PROMOTE SMALL AREA DATASETS AS A GOOD WAY TO RELEASE DATA WHILE MAINTAINING CONFIDENTIALITY

Custodians are rightly concerned with maintaining the privacy of the individuals that are the subject of their data. Many datasets are a collection of information pertaining to individuals, such that each person has a number of fields of information relating to them. One of these fields is often their address. We have found that data custodians are reluctant to release datasets that contain address information, or if they do release it, it's aggregated into such large areas (like states) that its spatial value is lost. One way to solve this problem, and still maintain confidentiality, is to aggregate addresses into small geographical areas.

For many government, community and commercial organisations, information aggregated into small areas can be used to make their work much more efficient and effective. As technology to process, display and analyse small area datasets advances, it becomes quicker and easier to use this information to add value to important decisions. Aggregating fields of information by grouping individuals who live or work in small geographical areas can protect the identity of those individuals. A minimum field (e.g. <15) can be set for small counts to protect the identification of individuals.

Custodians should consider small area aggregation as an option for their datasets and be able to access resources to help them publish in this manner.

---

### 2. IDENTIFY AND CLEARLY SIGNPOST THE PEOPLE AND RESOURCES WITHIN GOVERNMENT THAT CAN EDUCATE DATA CUSTODIANS ABOUT HOW TO RELEASE DATA

We have found that we are educating data custodians about how to release data in a way that is valuable to external organisations, but maintains the confidentiality of the data subjects. It would be great if this role existed inside government. It took us over a year to connect with the regional dataset team in the ABS, who can help data custodians with their releases, but may not have the resources to do so. As well as having resources to assist with data release, custodians need to be able to locate those resources easily. Contacts and resources for data custodians should be clearly signposted in obvious places like data.gov.au and its state equivalents. These resources should include clear and accurate information on legislation and policy around data release and should be public, so that internal and external agencies can have informed conversations.

---

### 3. IDENTIFY AND CLEARLY SIGNPOST THE PEOPLE WITHIN GOVERNMENT WHO CAN APPROVE OR ADVISE ON DATA RELEASE FOR CUSTODIANS

One of the most significant barriers to data release is the lack of clarity around governing legislation and policy. This is compounded by a cultural reluctance to share information. Agencies are risk averse around releasing data and lack confidence in what might happen to their programs and clients as a result of the data being used. A government authority that is familiar with the legislation and policy and can advise on release would solve this issue. This proposed authority would have a different role from that of the open data policy teams who advocate for the release of more data. The proposed

authority would need to understand the minutiae of the relevant laws and policies, and the technical processes of organizing and releasing data that can help custodians to comply.

---

#### 4. INCLUDE SOMEONE WITH TECHNICAL SKILLS IN OPEN DATA POLICY TEAMS

We have found that those in government advocating for open data and those in government who are custodians of data are not connected to each other, and that we are playing this connecting role from outside government. To date, the role of open data policy teams has focused on advocacy while the needs of data custodians in terms of making their data open are quite technical. Open data teams are doing a great job and we're seeing a lot of great policy released, but in order to implement these policies, the teams need to include or work closely with those with the requisite technical skills. Open data policy teams need to include people with the capability to actually help custodians open their data. Over time, this should include a cross-over of technical processes and knowledge between custodians and open data policy teams so that the process of opening data becomes more efficient and less costly.

---

#### 5. WORK WITH LOCAL GOVERNMENTS TO RETAIN DATA IN SMALLER AREAS

In several states, local governments are merging to create larger jurisdictions. We run the risk of losing richness in data categorised by local government area as part of this process. As the mergers progress, collectors and custodians should make efforts to maintain as much as possible the smaller data collection areas in their new systems. This data would be able to be aggregated to the new Local Government Area, but could still be split into the previous smaller councils for analysis.

---

#### 6. MAKE IT EASIER TO REQUEST DATASETS

It is currently very difficult to find the appropriate person within a government department to contact about a dataset and then figure out the correct process to request the release of that dataset. Most of the processes available assume that the data is being requested for a one-off use and ask us to detail the intended use of the data before it is released. We have found that there is usually no process for requesting that datasets be released publicly. It would be more efficient and transparent to have a single point of contact in each government department that can direct traffic to the right custodian or a publicly available list of data custodians with a description of their data. We envisage this role to be analogous to a freedom of information officer.

---

#### 7. ENCOURAGE CUSTODIANS TO REMOVE BARRIERS SUCH AS LOGINS TO SEE PARTICULAR DATASETS

Some custodians require users to log in to access their datasets, for example Transport for NSW <https://opendata.transport.nsw.gov.au/>. This means that when a data point is referenced, there are several more steps to checking its source, which means people are less likely to do it. To encourage evidence-based decisions, we need to allow decision-makers to quickly view raw data.

---

#### 8. CREATE A PROCESS TO STANDARDISE THE DEFINITIONS OF STATE DATA

The collections of the productivity commission have created standardisation for a lot of state data.

Commencing a dialogue about standardizing other data collections would be the first step in a long process of using the same definitions, collection timeframes and geographical categories across all state data.

Comparable state datasets would enable linkage projects, help states learn from each other and also make life easier for organisations that work across state borders.

## USERS OF DATA

When public data is made available, its value is often limited by the assumption that its audience is individual community members making personal decisions. The My School website is a good example of this ([myschool.edu.au](https://myschool.edu.au)). While this is an important audience, it's not the only one. At the other end of the scale, Transport for NSW (<https://opendata.transport.nsw.gov.au/>) has assumed that the primary audience for their data is app developers, which creates barriers for other users to access their datasets.

To maximise the value of data, it needs to be available or adaptable to the broader range of audiences that want to use it. Any publisher of data should check that it's easy to find and download their raw data, before tailoring it to a specific audience.

We take raw data and make it meaningful to organisations making location-based decisions for people in need. Users of our data platform include government departments and not-for-profits. These organisations are important consumers of public data and use it to make valuable decisions for services in our communities.

It's interesting that the scope of this Productivity Commission inquiry looks to "provide examples of public sector datasets that would provide high value to:

- the public sector,
- research sector,
- academics and
- the community."

Missing from the list are not-for-profits and commercial businesses, organisations that are increasingly delivering government services. These organisations could vastly improve their service planning and delivery if they had better access to public sector datasets.

## USES OF DATA

---

### EXAMPLES OF DECISIONS THAT COULD BE IMPROVED WITH BETTER ACCESS TO SPATIAL DATA

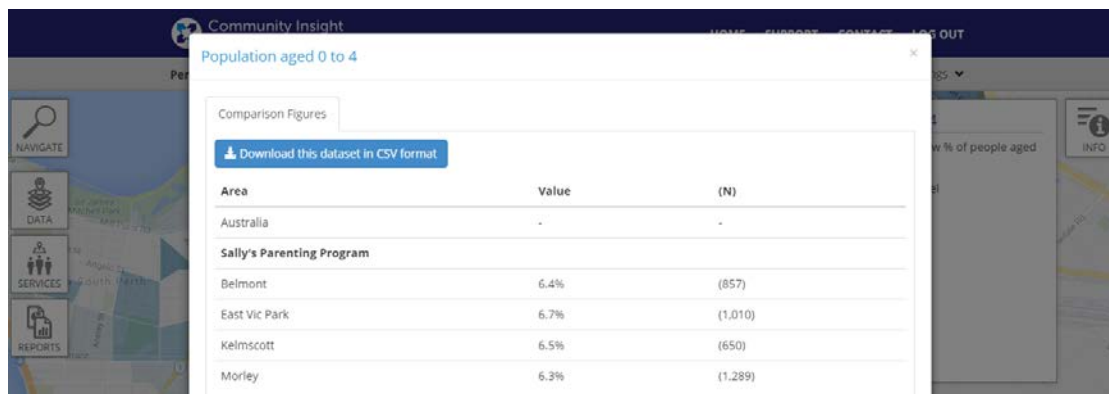
---

#### SARA – PROPOSING A PARENTING SUPPORT PROGRAM

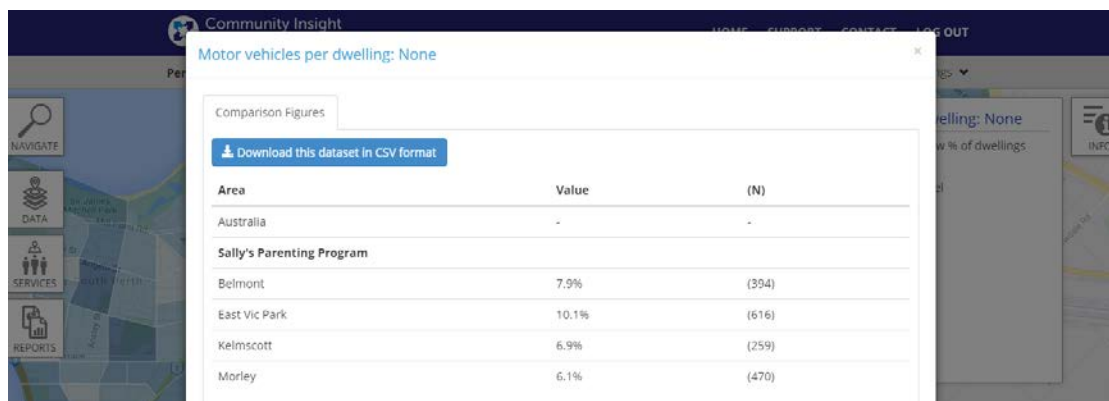
Sara is proposing a new parenting support program in response to a WA Government tender. The government wants to target young families by providing weekly support and connections to local services. Sara's organisation has a strong presence in Belmont, East Vic Park, Kelmscott and Morley, so she defines each of the areas in Community Insight Australia.

	Population aged 0 to 4	Working part time	Employed away from home	Unemployed and looking for work	Weekly household earnings	Weekly household expenses	Highest year of school completed	People with a Bachelor degree or higher	Attending University	Rented Total	No internet connection	One parent family	Persons with need for disability services	Motor vehicles per dwelling
<b>Sally's Parenting Program</b>														
Belmont	6.4	24.2	6.4	4.7	32.7	3.8	4.8	12.3	12.9	37.9	20.4	17.3	4.7	7.9
East Vic Park	6.7	27.3	5.8	5.2	31.7	4.0	3.9	19.8	27.1	43.0	18.0	15.5	4.4	10.1
Kelmscott	6.5	27.6	7.6	5.8	37.9	1.5	4.2	7.4	8.1	26.2	22.3	20.2	4.8	6.9
Morley	6.3	28.7	5.9	4.8	35.3	2.1	7.3	11.2	15.1	26.4	22.5	15.8	4.3	6.1

She can see that, of the four areas, East Vic Park has the highest percentage of children under 5. She uses the heat map to find out that there are 1010 children under 5.



She also discovers that there are 616 households that do not own a car.



She designs her program to cater to the community of East Vic Park and builds the cost of a mini-bus and children's car seats into her budget.

Sara can do all of this using data from the census, but she wishes she had access to **child protection information** for each of the areas under consideration so that she could target the area where a parenting support program could most help families with young children at risk.

## ARI – EVALUATING THE TENDER

Ari works for the WA Government and is evaluating the responses to the tender for parenting support programs mentioned in the above example. He makes a shortlist of the top proposals, and then defines the areas each proposes to work in. He downloads a report on each of these areas and judges each tender on whether it meets the needs of the community proposed. The parenting program is

part of a government-wide push to intervene earlier with children who develop anti-social and criminal behaviour. He wishes he could easily map the **home suburbs of the children appearing in children's court**, so that he could target communities where youth crime is a particular problem. If he'd had access to that data as he wrote the tender, he could have asked organisations to propose programs in those areas from the start.

---

## JANE - BUYING NEW PROPERTIES FOR COMMUNITY HOUSING

Jane is on the assets management team of a community housing organisation. They are doing a strategic review of properties they own, trying to decide where to sell and where to buy. She creates 'neighbourhoods' by drawing around the areas they currently own in and areas they are considering investing in. She uses the dashboard to compare the economic descriptors of people in these communities and their median house prices, which informs the team's discussion on the trade-offs between where they can afford to buy and where their tenants will thrive. Part of the work of the organisation is to support homeless people into housing. Jane wishes they had access to the areas where **specialist homelessness services** are in greatest demand, so that they could look for properties that didn't ask people to relocate too far away from their support networks.

---

## ADDITIONAL EXAMPLES OF HIGH VALUE DATASETS FOR THE PUBLIC SECTOR, RESEARCHERS, COMMUNITY ORGANISATIONS AND INDIVIDUALS

- **Contracts.** These must be machine-readable. While it is nice to see the signatures on the contracts, it's more useful to be able to scan the contracts for key words. Contracts are often very long and their use is maximized if users can navigate directly to the part of the contract that is relevant to them. For this to occur they must be machine-readable. Example: The Newpin Social Benefit Bond Deed of Implementation is a 90-page contract that is publicly available, but the PDF is not searchable (link to PDF: [http://www.community.nsw.gov.au/\\_data/assets/pdf\\_file/0005/328028/Newpin-SBB-UCB-FACS-Implementation-Agreement-as-at-30-June-2015.pdf](http://www.community.nsw.gov.au/_data/assets/pdf_file/0005/328028/Newpin-SBB-UCB-FACS-Implementation-Agreement-as-at-30-June-2015.pdf) available for download on this page: <http://www.community.nsw.gov.au/for-agencies-that-work-with-us/our-funding-programs/social-benefit-bonds/newpin-social-benefit-bond>).
- **Report on Government Services (ROGS)** (<http://www.pc.gov.au/research/ongoing/report-on-government-services>). The recent cuts to the scope of data reported in the 2016 Productivity Commission's Report on Government Services are severe and alarming. The change from 2015 to 2016 is drastic – 286 data tables in the 2015 Community Services chapters reduced to 5 in 2016. Particularly missed are the data on unit costs – they were really useful for state governments benchmarking the cost of their services and signposted who they should be learning from. The data made a very rich time series, which enabled analysis of change over time. We had hoped that the ROGS data would become more granular over time so that we could map it at a sub-state level - it now appears we can't access it at all!