

# Submission

## 2016 National Research Infrastructure Roadmap Capability Issues Paper

<b>Name</b>	Tom Kvan
<b>Title/role</b>	AURIN Director
<b>Organisation</b>	AURIN

### Questions

This submission is prepared by the AURIN facility on behalf of its user community. This submission has been framed around, and makes the case for, the following specific points:

1. Additional capability areas to be considered should include: **cities** and **indigenous health**.
2. Infrastructure funding should include support for the parallel processes to **build knowledge and skills across the education, government and industry sectors**. Shared training infrastructure and resources across NCRIS facilities will improve service provision across the scheme.
3. Infrastructure capabilities should include **cross-facility partnering for skills, tools and mechanisms**. Data capabilities such as AURIN are used to address emerging issues across domain areas such as urban settlements, health, geosciences, environment (energy and water use), all of which are critical to addressing future challenges and maintaining longitudinal data warehousing requires *increased upfront expenditure as an investment to enable increasing volumes over time as the repository rapidly grows in depth and breadth*.
4. **Urban data increasingly underpins research including in all disciplines**; tools or mechanisms are required to overcome barriers related to data standards, access, licensing, discovery and security.

### Question 1: Are there other capability areas that should be considered?

AURIN is also a contributor to a joint submission by TERN, RDS, AMMRF, Groundwater, APN, AuScope, PHRN, APPF and BPA NCRIS facilities. Our views regarding the broad questions raised as part of Roadmap development are addressed there. Within this submission we have focused on emerging trends related to the communities we support.

As argued in the joint submission we recommend **that cities, in all their complexity, are central to the future NCRIS Roadmap and constitute both a stand-alone capability area and provide a contextual basis for cross-domain integration of research activities**. Capabilities that integrate across disciplines, enabling them to collectively address emerging domain-transcendent issues such as urban settlements, health and environment (energy and water user) are essential, to addressing future research and

societal challenges. These have special value and need associated research infrastructure tailored to users (including both academics and government users).

Given this premise, there is a clear need to ensure that the other identified capabilities support such integration efforts. Thus whilst there is a generic need for “underpinning research infrastructure”, we believe that there are several targeted solutions that can be mapped onto different community needs and demands. For example, there is no single solution that would support urban research communities and high-energy physicists, however there may be some cross-cutting capabilities that allow both research communities to benefit. One such example may be, supporting the large scale aggregation of social media data and providing targeted analytics to underpin a range of fields and capabilities i.e. discovering data, cultures and communities etc.

We also believe that the capacity to accelerate research to improve indigenous health should be a priority. Currently AURIN is working with a number of project partners (the 6 year Department of Environment funded Clean Air and Urban Landscapes project; and with data journalists at the of Melbourne, and groups at the University of Queensland) to establish data related to indigenous issues. AURIN provides the technology to act as a clearing house for these datasets. Prioritising additional enhancements to fulfil specialised indigenous data collection and sharing requirements would provide a significant research capability in this area.

**Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.**

The governance of a capability must be aligned to the nature of the capability. AURIN is a social science data access portal and thus raises questions distinct from those of science hardware.

We start by noting that the NCRIS issues paper describes an **Ideal Research Data System** on page 48 (reproduced as Figure 1) that accurately reflects the AURIN Data Model. Data and metadata are collected from myriad sources including administrative systems, sensor networks, research outputs and on-ground surveys. The data is programmatically made accessible to the AURIN Portal, which provides secure access to users according to the specific license conditions of the data provider and the user privileges. Users can search for and discover data sets using the metadata provided (this metadata is provided by the data agencies); this searching can be geospatially bound, e.g. unemployment statistics for Greater Melbourne. The Portal environment also contains over 100 analytical tools for analysis and collaboration through workflows leveraging the Cloud and thereby enabling data combination, transformation and translation for further dissemination and reuse (subject to data under licence conditions). AURIN extends this ideal research data system through enhancing capabilities for the dissemination and verification of published research data, and more importantly providing the methodology used for data analysis. AURIN provides the ability to link actual data with the analysis (workflow) used in a given research paper. This ability to verify and validate research results is increasingly important and has thus far been inadequately supported.

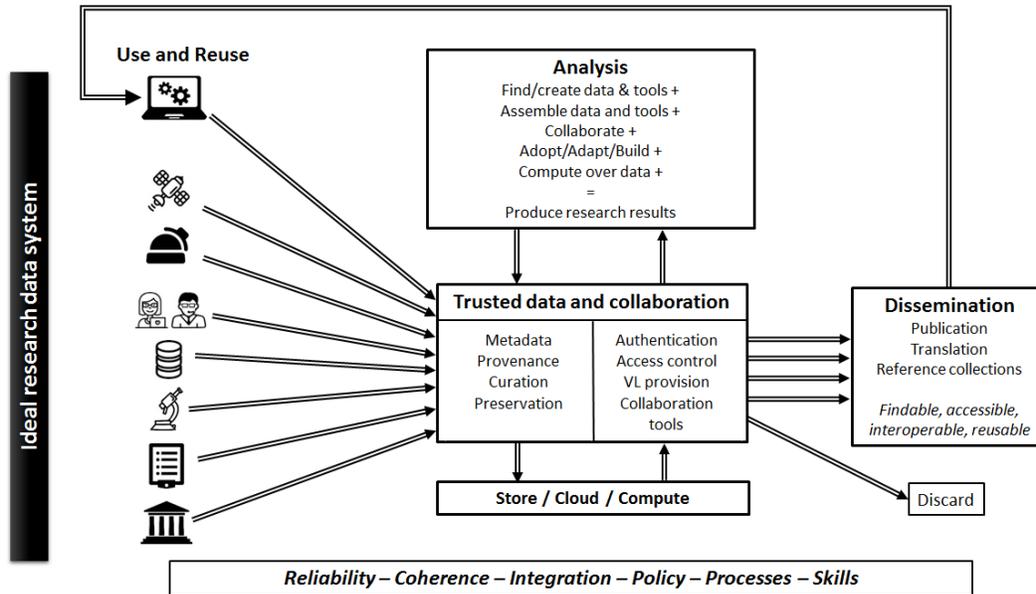


Figure 1. Conceptual model of idealised NCRIS data infrastructure and the model already realised by AURIN.

Creation of the AURIN environment, or any environment similar to Figure 1, involves considerable systems development and implementation. Less obvious, but no less important, is the significant contribution made through organisational and procedural developments. Technology development alone is insufficient; relationships and trust developed over time with key data providers are crucial to maximising the value of this type of infrastructure and governance must support this effort. AURIN has over 70 such relationships with key data-rich agencies crossing the urban settlements. Overcoming barriers and supporting long-term sustainable relationships with such agencies is therefore a central element of this submission. Indeed, the major procedural barriers to effective data infrastructure were identified in the joint submission (with TERN, RDS, AMMRF, Groundwater, APN, AuScope, PHRN, APPF and BPA). These barriers, which vary slightly depending on whether the infrastructure is physical or virtual, are related to:

- intellectual cost;
- financial cost;
- institutional arrangements;
- geographical differences, and
- invisibility of sources.

The governance characteristics outlined in the issues paper are broadly appropriate, however, we note aspects of **board structure and independence**, further rigour around **intellectual property strategy** and, in particular, **security and certainty of funding** without which trust cannot be established.

Optimal governance includes the security to plan for the future. Five year funding cycles would facilitate staff skills development, testing of ideas, and the innovation and translation of research findings into industry. Current short timelines severely restrict the infrastructure's development potential. AURIN has

first-hand experience of the consequences of the current uncertain funding model. Highly trained technical staff with experiences at the leading edge of big data systems have left to join Amazon in Germany, investment banks in London and the National Broadband Network.

AURIN supports a governance model at project level that provides a combination of strong leadership and expert ability at Chief Executive Level with an independent board informed by governance principles for not-for-profit companies. Board members are key proponents for the infrastructure and their independence augments the expertise required for infrastructure delivery, it shields infrastructure from institutional politics and fosters collaboration and decision making across a broad national community of collaborators and contributors. Similarly, engaging independent expert legal and business advisors where funding/co-investment permits, could preserve the independent direction of infrastructure projects and strengthen engagement with the business community and support for national investment priorities.

**Question 3: Should national research infrastructure investment assist with access to international facilities?**

Governments across the world are focussing on cities as economic drivers and looking for greater efficiencies, in particular their capacity to foster innovation and deliver a higher quality of life for the citizens. Policymaking and infrastructure investments in cities are increasingly supported by an evidence base and proposed strategies tested by scenario modelling. Infrastructures such as AURIN are increasingly being funded and cross-nation comparators being used to guide decision-making.

We observe that the value of international linkage is more likely to be through partnerships and technology/knowledge transfer from Australia to other nations. For example, AURIN has a Memorandum of Understanding with the University of Glasgow – Big Urban Data Centre and have been invited to showcase its technical developments to a number of international organisations in New Zealand, Korea, the Philippines, PNG and Fiji. Discussions are well advanced with a range of universities in Singapore on Smart City analytics and potential for joint research efforts and collaborations more generally. Our workbench tools are currently being tested and/or used in Pakistan, Egypt, USA, UK, South Africa, Iran, Germany, Saudi Arabia and Turkey. Partnerships such as these enhance Australia's reputation and provide an opportunity to be world leading in this area.

Work is ongoing to support access to international researchers through the Australian Access Federation and involvement in international federations. The technical aspects of such international collaboration are well in hand, however Australian data providers are cautious about allowing international use of their data especially when the data has been provided for Australian academics.

**Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?**

As outlined in the joint submission, if international facilities can be accessed more cheaply or efficiently than those that may exist (or not exist) in Australia, then we should encourage those options. It is noted that with the specific focus on cities, it would be highly beneficial to support comparative analyses with

other global cities. At present such comparisons are difficult or impossible to undertake due to limitations on access to and use of comparable data and systems accessible overseas.

**Question 5: Should research workforce skills be considered a research infrastructure issue?**

Three major factors are driving the demand for urban data and associated skills. First, educators are required to teach students how to build a holistic understanding of the world using technology and analytics. Second, policy makers are looking to evidence on which to develop and test policy initiatives; a data foundation is fundamental to this work. Third, managers of urban systems are looking to gain advantages inherent to the smart cities approach, again this relies upon good data to deliver a better quality of life for citizens. More broadly,

1. For educators developing student skills in the application of technology as identified as a critical need for Australia (Australian Council for Educational Research, 2010; 2015; Office of the Chief Scientist, 2014). AURIN is collaborating with the Geography Teachers Association Victoria (GTAV) to build skills in the application of urban analytics, spatial technology, mapping and the use of information within the context of contemporary issues and data comparability. The GTAV recognises that AURIN is an important component in of Science, Technology, Engineering and Mathematics (STEM) and fosters the development and application of distinctive STEM skills which are important in modern geography.
2. Policy makers are looking to evidence on which to develop and test policy initiatives; both knowledge to interpret data and skills to handle are fundamental to this work. Additionally, managers of urban systems are looking to gain advantages inherent in the smart cities approach, again this relies upon good data to deliver a better quality of life for citizens. More broadly, industry and researchers in Australia increasingly understand the power of data, in particular spatial data, this in turn creates a growing demand for skills and talent to best use the data. Where research infrastructure is focused in a particular domain, skills development is restricted to researchers within that domain. Conversely, infrastructure that supports users, supported by appropriate training, in multiple domains provides a greater boost to national skill levels. Currently AURIN has a user base of over 5,100. The facility regularly conducts training across a diverse range of discipline areas and is able to introduce the power of location based data and spatial analytics to transform research models across multiple disciplines. This allows the development of skilled communities from citizen scientists, through private enterprise and on to world class researchers.
3. Both infrastructure-developer skills and infrastructure-user skills should be considered as priority research infrastructure issues. The technical staff that implemented, delivered and supported the AURIN portal and associated infrastructure are highly trained and acquired additional skills and experience throughout the project. This project knowledge takes significant time to develop and strategies to maximise retention of these skills is essential. In addition, AURIN has approached systems development in parallel with user training and skills development. This ensures the highest uptake of the research infrastructure and maximises its impact and worth. This emphasises the importance of weighting training and outreach equally to building a system; as without skilled users the infrastructure system will not reach its full potential as a national asset. Furthermore

many NCRIS infrastructure platforms run similar training and skills development programs, some efficiencies may be gained by considering integrated approaches to training and knowledge transfer.

**Question 6: How can national research infrastructure assist in training and skills development?**

To ensure the confident uptake of AURIN assistance could include promotion of the infrastructure in professional learning streams including schools, universities and the public and private sector. Figure 2 provides a simplified diagram of the approach. The AURIN system is underpinned by access to authoritative datasets that are typically located with data custodians. This data is accessed, harmonised and analysed by researchers via the AURIN portal. This live access to definitive data is key since establishing a data warehouse with copies or snapshots of data would rapidly become obsolete. The portal includes built in capacity to support users with different skill levels and needs ie. access to raw data, processed and combined data for derivative information, and spatial and non-spatial analytical tools capable of extracting knowledge for a rich variety of research areas. Researchers, functioning at different skill levels, contribute to the reuse and refinement of data and can achieve a collective benefit. Importantly, urban settlements are inherently cross-disciplinary reflecting the multiple overlapping associated challenges, such as housing, income, transport, crime and health. Research infrastructure must enable cross-disciplinary collaborations and not stifle them. Hitherto, data and research silos were the norm in urban and built environment contexts. Future NCRIS efforts must ensure that multi- and inter-disciplinary research efforts can be supported through targeted infrastructures.

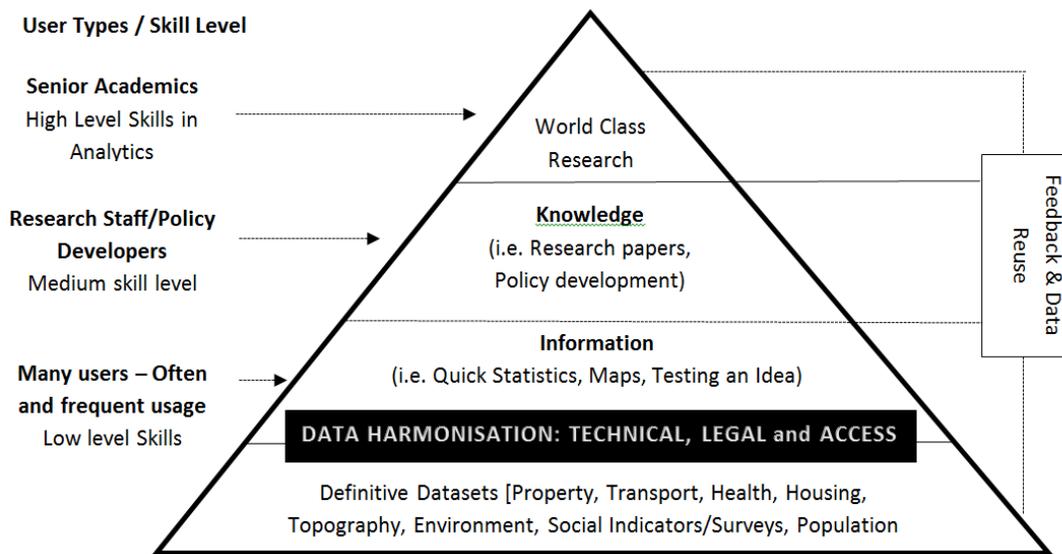


Figure 2: Relationship between a data and the overall objective of AURIN – to support the continuum of urban research across a multitude of urban research areas.

Data created through a number of the NCRIS facilities is now provided through many channels (portals, APIs, apps) and therefore training mechanisms should also be available in diverse forms. Whilst we propose to continue and extend our training programs, we see further fostering of on-line training support as essential to the skills development of an ever wider user community. Skills development is not just essential for systems developers and managers but, in the case of data-focused research infrastructure, for the end users also.

AURIN infrastructure provides a rich source of data that is used for educating the next generation of software engineers. The AURIN platform forms a cornerstone in several courses that are ongoing at the University of Melbourne, e.g. Cluster and Cloud Computing. Students are exposed to the big data technologies that have been used and underpin the AURIN platform. Obtaining hands-on experiences in overcoming data heterogeneity or big data processing on the Cloud is in great demand and many of the students that have completed this course have since joined major organisations, e.g. Google. The course has now been running for 4 years and trained over 500 typically-Masters-level students.

The AURIN platform also provides a rich source for other software engineering activities. The openAPI that has been developed for AURIN has formed the basis for numerous mobile applications completed by Masters-level students. Over 100 Masters-level dissertations have been completed on a range of topics using the AURIN data. Examples of some of these are available at [www.eresearch.unimelb.edu.au](http://www.eresearch.unimelb.edu.au).

**Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?**

AURIN was developed in a collaboration of research institutions across Australia. However, long term maintenance rests with the host institution. One of the challenges is the cost associated with maintaining data supplies, hardware and software coupled with costs required to develop the infrastructure to ingest new data types. As a result the relevance of the system is eroded over the long-term.

We also strongly believe that research institutions have a role in supporting researchers outside the traditional NCRIS focus of scientists and engineers, such as architects, planners, social scientists, linguists, criminologists etc, to take advantage of emerging opportunities. In the big data era, data now comes through many channels (portals, APIs, apps linked to WFS, WMS, CWS and other distribution protocols) and training mechanisms should also be available in diverse forms to different potentially less tech-savvy communities to ensure that they can also benefit. The AURIN team has given many presentations and hands on training demonstrations: among them, big data analytics in the financial domain; data science in the Earth Sciences; and social media analytics in the Humanities.

**Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?**

The AURIN platform has two parts data and software. In principle, the software developed national research infrastructure should be open and accessible for little or no cost and data should be available with ease to researchers, with commercial access provided on a cost recovery basis.

One of the barriers to high value data is the cost of private sector data which is currently limiting high impact research. We note that we have many years of experience in delivering a platform offering multiple tiers of access to different communities. The technical challenges in delivering this are often far outweighed by the non-technical issues that surround data access, e.g. licensing agreements between legal entities.

By default the AURIN model is to provide access to all data for all academic researchers wherever possible. There are however a few data sets, e.g. ABR where restrictions are imposed. Going forward we believe that this model is essential to the success of the platform and a charge/cost recovery model would be difficult for the community to accept/support.

**Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?**

Clearly, if a facility is not meeting agreed KPIs and/or it is not being used then it should be defunded and/or decommissioned. In decommissioning a data infrastructure such as AURIN, a number of specific issues are raised. When any decommissioning, it is important to assess the assets register as it may be possible to utilise infrastructure, software and data within other capabilities. Should data be retained as a historic snapshot of the moment of decommissioning or should all access be removed. We note too that this decommissioning can have impacts on others, e.g. data providers, while removing the foundation for future research. Defunding or decommissioning NCRIS-funded infrastructure would require in-depth analysis of the business case for and against with current and future scenarios taken into account; Return On Investment (ROI); user satisfaction; impact assessment (not dissimilar to the qualitative assessment currently being considered for research impact).

It is noted that AURIN has numerous KPIs that have been systematically followed throughout the lifetime of the project. These have been independently assessed by both national and international expert groups, and their guidance and feedback has informed the evolution of the platform. We strongly believe that all NCRIS capabilities should have associated KPIs that are used to assess their activities. Arguably the strongest KPI is the usage and uptake by the community, and especially by the communities that have no vested interest in the capability, e.g. they receive no funding from the capability. In the case of AURIN, we have a significant portion of the urban research community and a growing user base in government and industry. We aim to continue to grow these communities.

**Question 10: What financing models should the Government consider to support investment in national research infrastructure?**

Funding models to date have failed us by being short-term focussed and not admitting alternative funding streams to supplement needs. If the Australian Government wishes to remain globally competitive in education and research, then ongoing funding is essential. Although the models of funding through EIF and NCRIS have established world leading infrastructures that are in many cases the envy of the world. Future funding arrangements might consider leveraging State-based funding to complement the federal investments and allowing for industry funds to be sourced, either as capital

investments for additional capabilities or through ‘fee for service’ activities. Obviously the implications here are significant in terms of governance, security of access and future developments.

On a more pragmatic matter related to data focused NCRIS capabilities, significant investments are made on creating data, but little monies spent on optimal ways to re-use or re-purpose such data. There are huge cost benefits that could be achieved if the infrastructure for data sharing is established and supported as an ongoing activity and becomes “the norm” as opposed to the current model of major NCRIS investments to establish “the unique platform”. To achieve this, Government departments publishing data should be funding the structuring of the data so that the costs are defrayed from the single source funding bodies such as the ARC, NHMRC and RDCs. These groups should be encouraged to support the NCRIS facilities and there should be mechanisms in national research funding schemes (such as ARC) for grants proposals to identify cost and therefore income to support on-going infrastructure access and ongoing development costs.

**Question 11: When should capabilities be expected to address standard and accreditation requirements?**

As in our Recommendation 2, NCRIS capabilities should comply wherever possible with national and international standards. In AURIN the relevant standards include the Australian New Zealand Standard Information Code (ANZSIC) which relates to classification of themes, ISO 8000 which relates to data quality, and ISO 19115-1:2014 in geographic information and metadata. At the same time, it would greatly assist our operations if our data providers also followed international or ANZ standards to ensure complementarity across jurisdictions, e.g. land-use zoning and crime statistics are prime examples of ‘failures’.

Additional standards may also be required to address the following issues:

- *Third party data collection* – Data collection is often outsourced to consultancies and is regularly presented in paper rather than digital format, which severely restricts use and reuse of public data.
- *Licensing* - Uncertainty surrounding ownership of data and how it can be used is a key barrier for the use and reuse of data. Data licensing plays an important role in informing users how different datasets can be used, which in turn enables the organisations and individuals that created the data to be appropriately attributed and acknowledged.
- *Privacy and ethics* – A number of questions arise in this space. Whose data is it if it exists in the public domain, such as social media? Similarly, health records, business registry data etc. Clarity is required along with widely accepted processes for generalising data to maintain confidentiality. Resolving these issues will provide a clear sense of what is legally appropriate in relation to data and information derived from confidential data, and for instance, provide better solutions for capabilities as identified in section 5.3.2 of the issues paper (Indigenous research platforms).
- *Non adherence to standards* - In general we believe that NCRIS capabilities should comply with international standards. As mentioned above, it would greatly assist our operations if our data providers also followed international or ANZ standards to ensure complementarity across

jurisdictions (land use zoning and crime statistics are prime examples of incompatibility between the States).

- *Data invisibility* - Often data is intimately linked to a single individual researcher, or a small team, and it is very hard to find that source. A national “find an expert” facility would help other researchers to locate these ‘resources’ especially when the research is cross disciplinary and so awareness does not emerge via the usual channels (e.g. conferences, publications).
- *Security concerns* - There’s a patchwork quilt that exists across the whole NCRIS space. There should be a “centre” that offers security advice with leading experts and are funded to provide it for all sectors of community as appropriate – e.g. Australian Access Federation (AAF) is only available for academic institutions.

AURIN has been active in several of these areas but they need to be generally recognised and the quest for solutions supported.

**Question 12: Are there international or global models that represent best practice for national research infrastructure that could be considered?**

There are many major efforts that have taken place in the urban and social sciences that reflect best practice research infrastructure. In the UK the UK Data Archives ([www.data-archive.ac.uk](http://www.data-archive.ac.uk)) is recognised as a major infrastructure provider offering key services to the social science community. This has been established over many years through grants from the UK Economic and Social Research Council (ESRC). There have also been several efforts to harmonise social science data archives, e.g. the Consortium of European Social Science Data Archives (CESSDA – [www.cessda.net](http://www.cessda.net)). There are also many other activities that are occurring through the Horizon 2020 program (see for example <https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/calls/h2020-scc-2016-2017.html> and <https://ec.europa.eu/inea/en/horizon-2020/smart-cities-communities>). The Horizon 2020 program has €87billion (AUD127b) of investment, hence aligning with the future NCRIS efforts would be highly desirable.

It is important to note that these are targeted services to particular communities (in this case social sciences) and not generic solutions across research domains. This model is important to understand since ignoring the targeted community needs is a risk. From past experience of working in a range of projects have been decommissioned, we refer to the National Grid Service (renamed as the ‘National e-Infrastructure Service’) and the Enabling Grids for e-Science, which became the European Grid Infrastructure. Both of these efforts were primarily generic infrastructure activities, i.e. they provided generic solutions focused predominantly on computationally bounded problems. The vast majority of researchers have data issues and are not computationally bound. AURIN is much more of a data-driven solution that maps directly onto research needs. That said, we do have computational services, but these are targeted to the community needs directly, e.g. walkability. Our offerings are better aligned with the communities they serve and are not in the “build it and they will come” nature of National Grid Service (NGS) or European Grid Infrastructure (EGI).

**Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?**

This is addressed in the joint submission. We note here that there have been several examples of decommissioning/defunding, e.g. NGS and EGI identified above that might be worthy of reflection of why they were decommissioned.

**Question 14: Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?**

Current funding models do not recognise the full cost of developing and maintaining infrastructures nor do they facilitate additional funding streams to support ongoing costs. There need to be mechanisms to benefit from both State-based investments and federal activities that rely on the infrastructure, as well as for commercial and industry users to gain access. One alternative model that should be actively explored is through engaging in larger-scale international efforts. Thus it is possible through NHMRC to be involved in Horizon 2020 efforts. A specific set of projects/calls for funding is applicable here. Ideally this model could/should be extended to other non-health domains.

The FARR Institute ([www.farrinstitute.org/](http://www.farrinstitute.org/)) brings 21 academic institutions together from across the UK to analyse data to better understand the health of patients and populations using ‘big data’. FARR is supported by a consortium of ten funders, led by the Medical Research Council to the value of £37.5m (AUD 65m). Similarly, the Administrative Data Research Network (ADRN) helps accredited researchers carry out social and economic research using linked, de-identified administrative data from government organisations, in a secure environment for the researcher to work in. Both FARR and ADRN are funded by the Economic and Social Research council (ESRC), UK’s £200m (AUD 348), the public sector funder of social science ([https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-adrnboardcompositeslidepac\\_tcm97-43883.pdf](https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-adrnboardcompositeslidepac_tcm97-43883.pdf)).

## **Understanding Cultures and Communities**

**Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?**

We believe we need to think beyond simply understanding ‘Cultures and Communities’ into building future scenarios based on a solid evidence base. The resources available through AURIN can assist in designing a future that is socially just and environmentally sustainable.

One of the key aspects of understanding culture and communities is the inter-relatedness of data in spatial settings. There were many challenges facing AURIN at the project outset in convincing the community that a one-stop shop would meet their urban research needs and demands. We believe that we have successfully met all of those challenges and established a vibrant community of diverse users. The future landscape for Understanding Cultures and Communities must encompass this diversity of interest. It is quite tempting to put silos together on aspects of urban settlements, e.g. focusing solely on

health, or solely on transport etc. Cities are complex systems where multiple researchers and multiple subsystems need to be brought together to understand particular urban phenomenon.

Although the work of AURIN was contextualised in major cities, there are a multitude of rural, regional as well as urban data sets that are available in the platform. The distinction can be somewhat arbitrary. Future NCRIS initiatives in Understanding Cultures and Communities should embrace all spatial levels and contexts.

One of the major efforts planned for the Agility part of AURIN was introduction of social media into the platform. Whilst many official organisations hold definitive data, e.g. the Census, many more researchers are interested in the events that are happening now that impact on the cities in which we live. Social media offers one data resource that can capture the heartbeat of cities. At present many groups across Australia are undertaking smaller scale social media analytics, however ideally there should be a national platform offering major data holdings with multiple analytics as a service to the community. This should ideally have official relationships with these platforms, e.g. Twitter, Facebook. Having this negotiated through the Government can help unlock more data than any individual or any project could achieve.

**AURIN's unique position as a national research infrastructure illustrates how value can translate across a multitude of national research areas of direct importance to society**

Pressing social issues, critical economic drivers and environmental changes over the next decade will see unprecedented demands placed on Australian cities and regions. As a result, Australian research communities need to plan for, and respond rapidly to, the challenges ahead. Access to data to inform public policy and innovation is critical. This is particularly important as the nation plans for population growth and transitions to the digital economy, and will address issues such as: how population should be distributed across space, demand for skills and jobs, the types of urban environments in which people will live, and how the nation can achieve the sustainable development of its cities and surrounds in response to changing climate, demographics and communication technologies and consequent life styles and work patterns.

AURIN is giving Australia both national data advantage and significant operational efficiencies. We estimate that an average research project can save approximately 500 hours that would otherwise be used in sourcing, vetting, negotiating and accessing data. The savings in legal costs are also significant. If, for example, the average AURIN user requires three data sets for their research, this represents approximately 30-40 hours of legal negotiations per 'high value' dataset including discussion of and review of terms and conditions. AURIN has already negotiated and secured access to 1,600 datasets on behalf of our users, including a number of 'high value' datasets not available on open-data portals (average saving estimated at \$10,000 per user).

Across Australia AURIN is being used consistently to understand complexities within cities and anticipate future challenges and where appropriate establish pre-emptive data-driven policies. Over 5,100 individuals have registered with AURIN and 30% of these are engaged users. It is important to note that the use of the infrastructure and data available has gone far beyond the original focus within the

domain of human settlements. The following examples demonstrate how AURIN has been used to translate data into research outputs and insightful policy development:

- a) The National Health Services Directory (NHSD) provides an authoritative record on the locations and details of health service providers nationally- this helps form policy around geographic disadvantage, and allocation of services with regard to ageing populations or communities with intensive needs – both urban and regional. Dr. Melanie Davern outlines the importance of this data for her research

*The Directory provides an amazing source of geocoded health services data. We are currently using the data on General Practitioners located across Victoria with special interest in the ability to speak a language other than English for our research investigating health services provision and planning for humanitarian entrants to Victoria. This is innovative and novel research partnered by the Department of Health and Human Services, Department of Premier and Cabinet, and many service providers and community organisations. Thanks again for the support and partnership of AURIN for sourcing these data."*

- b) Assisting the Australian Business Register (ABR) to provide their geocoded data to all levels of government. This has been described by the ABR as *"...a pioneer in enabling access to the ABR information through a 3rd Party provider. Access to the ABR information via AURIN enables eligible government agency researchers, policy analysts and decision makers access to a contemporary platform it also enables the ABR information to be provided as special aggregated data for all university researchers. Provision of this type of aggregated information is also a first (outside of the Australian Bureau of Statistics) and will support evidence-based policy and decision-making research such as service planning, disaster analysis and response (e.g. fire/flood mapping), productivity, congestion and urban resilience.*

*We are delighted to be part of this initiative and look forward to supporting your agency and the innovation it enables for developing Australia's future."* Sandee Harris (Australian Taxation Office). Like the NHSD, these data are regularly updated to enable longitudinal change mapping.

- c) The AURIN platform is supporting research communities in a range of health disciplines. For example the national paediatric type-1 diabetes platform has adopted AURIN to map the distribution of patients and combine this data with socio-economic data to better understand the environmental impact of diabetes; the national spinal cord injury platform has adopted AURIN to better understand the issues and challenges of accessibility for spinal injured patients in/around cities; and cancer researchers at the Peter MacCallum Cancer Centre are using AURIN with mobile air quality sensors to assess the impact of air quality on a range of lung-disorders (e.g. cancer, chronic obstructive pulmonary disease, asthma etc).
- d) The AURIN platform provides a model of how agencies (local, state, Federal government), industry and academia can securely share data in an on-line programmatic manner. This approach is shaping new innovation for businesses and savings. AURIN's collaboration with Melbourne Water demonstrates the relationship: *"Melbourne Water has been supporting AURIN since 2014, to share data for research purposes. This has reduced the number of requests we receive and supports Melbourne Water's open data policy."* (Kim Rennie, GIS Analyst) Monitoring of the data downloads shows that savings to Melbourne Water is on average 50 hours work per

month. Melbourne Water is now also benefiting from the innovations developed from the use of its data within research, this includes algorithms for the automated extraction of urban features. The relationship also promotes AURIN within other water utilities as they investigate how to fulfil Open Data Policy objectives.

**Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?**

As noted earlier, governments across the world are developing 'smart city' capabilities to enhance competitiveness and improve quality of life experiences. Data underpins these plans. For example, the Indian *100 Smart Cities* initiative (<http://smartcities.gov.in/>) aligns with Australian efforts to promote improve quality of life, drive economic growth, create employment and maintain or regenerate clean and sustainable environments by best applying new technologies. This initiative has a total of ₹980b crore (AUD 19b) approved for development of 100 smart cities and rejuvenation of 500 others across India.

AURIN has been supporting this work locally. The Rockefeller foundation's *100 Resilient Cities program* provides 100 member cities with \$1m funding to hire a Chief Resilience Officer, and access to over US\$200m (AUD260m) in Services, and critical support to develop a City Resilience Strategy. This is a global initiative in which Australia can demonstrate leadership. The City of Melbourne is a member city and is participating in this program and the advantage has been described by Toby Kent, Melbourne's Chief Resilience Officer, "AURIN has been central to our work, both in providing initial information to support our project scoping to produce our Preliminary Resilience Assessment, published in June 2015. Since then, AURIN has supported 'Accelerated Design Fora', which involved cross-sector collaboration to rapidly identify resilience-building opportunities..." Data and maps from AURIN are prominent within the Resilient Melbourne strategy, released on the 12 May 2016. Building on this success discussions are currently taking place for AURIN to become a platform for the Rockefeller 100 Resilient Cities Initiative globally.

There are many emerging projects in Australia that are moving and shaping the direction of AURIN and showing how it can be adapted to support new communities, e.g. the \$8.8m Department of the Environment funded Clean Air and Urban Landscapes ([www.nespurban.edu.au](http://www.nespurban.edu.au)). Multiple new data sets have been included – demonstrating the flexibility and power of the platform. Such communities are keen to explore the long-term effects and impact of policies. Air quality and global warming, and the lifestyles we live, require accurate measurements to help inform future decisions on urban settlements. With appropriate infrastructure, Australian industry can benefit and position itself as a global leader in service provision for planning and decision-making.

**Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?**

By understanding communities and cultures, we can establish an evidence base for forward-looking development of scenarios for policy decisions and major investments. Underpinning this evidence must be a solid and reliable infrastructure of shared data. This is recognised as a significant challenge by

jurisdictions at all levels of government. These data are then the platform for many other areas of research in which their discipline specific subject matters are contextualised. Thus, in the subject matter of data, communities and cultures should not be considered a discrete research area but an underlying platform for national research capabilities. There are many studies that rely on authoritative and longitudinal metrics (See [Australian Health Services Research Institute \(AHSRI\) ahsri.uow.edu.au](http://ahsri.uow.edu.au)) to form the evidence base for which the long-term well-being of communities can be measured. Resolving privacy, sharing and cost issues are required from the outset in order for the initial investment in warehouses of observed data to be best designed for the long term.

## **Data for Research and Discoverability**

### **Question 33 Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?**

The future for data will include new data types - often involving very high volume, and/or high velocity data. These big data may come from various sensors, human sources (socially connected data) or from within public administration. Concurrently it is necessary to develop smart analytics capable of modelling and analysing such data to improve prediction and reduce uncertainty. These activities require world-class highly specialised skills and systems through which they can be effectively applied in multi-disciplinary contexts. Using these systems constitutes a particular complementary skill.

As an example, the AURIN portal supports linkage between health data and non-health data. AURIN's high value geospatial datasets coupled with specialised skills of analysis and interpretation represent an untapped opportunity to investigate the effects of social disparity and other socioeconomic factors on health.

Almost certainly there are lessons from international epidemiological studies that would be beneficial for AURIN to extend capabilities in the provision of big medical data. Big clinical and medical data is limited when used in isolation. However, linkage with social, geographical and environmental datasets enables very powerful connections which open up enormous possibilities. AURIN has much of the required social and environment data infrastructure in place already. Removing access impediments to government held health data would be an important first step in extending these powerful linkages. It is noted that AURIN is already being used in large-scale linkage studies with health researchers, e.g. the \$12m Environmental Determinants of Type-1 Diabetes study ([www.endia.org.au](http://www.endia.org.au)) is one major example where environmental data on the patients is used to better understand why some children will develop Type-1 Diabetes.

In the epidemiological context keeping historical time series data is also particularly important as it aids in understanding temporal relationships. There are a number of international sources of historical demographic data – e.g. [www.ipums.org](http://www.ipums.org) – with which AURIN can explore partnership arrangements.

A similar scenario can be developed in the area of Environment and Natural Resource Management. The capability to enable the collection and integration of datasets covering the range of human activities by those working, investing and living in urban areas is still required, with governments at all levels demonstrating renewed policy focus on cities and sustainability. eResearch tools need to be augmented

to link environmental and geoscientific datasets to form a holistic picture of urban settlements. The integration of societal observation data, along with environmental observational data (e.g. air quality), will initiate a new paradigm of urban research, enabling a more fulsome research effort on the design, operation, sustainability and liveability of urban environments and the impact that changes are having. In particular, we seek to extend traditional time-sliced data sources with real-time observational sources such as sensor networks and citizen sensors. (See our Recommendation 3)

**Question 34: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?**

The [European Commission's INSPIRE](http://inspire.ec.europa.eu/) directive is charged with aligning infrastructure for spatial information across Europe (<http://inspire.ec.europa.eu/>) and has been funded €100m (AUD146m) annually for 10 years. INSPIRE will reach completion by 2019 and supports the Horizon 2020 initiative. INSPIRE is a very similar project to AURIN and provides an important case study for national eInfrastructure implementation.

**Question 35: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?**

There is a clear need to listen to, and consult with researchers during the discovery, design and implementation of any infrastructure for collecting and managing data for research and discoverability. The vast majority of researchers have their own ways of discovering and accessing data. Often this is domain-specific or through generic web capabilities such as Google. AURIN has been successful in fostering discoverability because it has listened to users. It is essential that the future Data for Research and Discoverability capability reflects what researchers want and need, and is willing to provide for others through the infrastructure. Often researchers and agencies are acutely aware of the sensitivities and importance of their data. Systems that can provide secure access and use are essential and should be a cornerstone of any discoverability solutions. This in turn often requires that auditing and accounting systems are in place to ensure that access and use is being delivered in accordance with the data provider's needs and demands. AURIN has delivered such fully worked solutions to numerous data providers and researchers with sensitive data, e.g. ABR, PSMA, APM, VicHealth, WA Health.

There are many organisations and agencies moving towards open data at federal, state and city levels. The ability to simply use such data, e.g. making it accessible through AURIN, often requires a license. There should be standard license models that all open-data.gov agencies sign up to. This would make use and repurposing of data far easier.