The Australian Urban Research Infrastructure Network (AURIN) Initiative: A Platform Offering Data and Tools for Urban and Built Environment Researchers across Australia

Robert J Stimson
The University of Melbourne, Victoria, Australia

Abstract

The Australian Urban Research Infrastructure Network (AURIN) (www.aurin.org.au) is a $20 million project funded under the Australian Government's Super Science initiative. Its objective is to develop and implement an e-Infrastructure providing researchers access to comprehensive sets of diverse data from multiple sources with a capability to integrate and analyse those data. Researchers will be provided e-Research services enabling them to interrogate those data using a 'tool-kit' of analytical, modeling, simulation and visualization procedures. Developing the AURIN research infrastructure will involve close collaboration with other existing research infrastructure initiatives and with existing research institutions and public and private agencies. This paper outlines the AURIN project, its focus and the capabilities it will provide to researchers.

BACKGROUND

As part of the Australian Government's EIF Super Science initiative, the Australian Urban Research Infrastructure Network (AURIN) project (www.aurin.org.au) received $20 million for the '*establishment of facilities to enhance the understanding of urban resource use and managemen*t'. The government selected the University of Melbourne as the Lead Agent for the project, and a *Funding Agreement* was signed in June 2010.

The project will run until the end of June 2014. It aims to provide urban and built environment researchers with access to diverse sources of data, data integration and analytical capabilities, and tools for processing and storing those data. This e-Infrastructure will help facilitate research activity that will enhance our understanding of key issues relating to the urban components of Australia's national settlement system with the potential to provide an improved evidence base to inform public policy and inform business operations. This is particularly important as the nation debates the implications of population growth, how that will be distributed across space, the types of urban environments in which our diverse peoples will live, and how the nation can achieve the sustainable development of its urban settlements.

The intensions of the AURIN project were outlined in a document prepared for the Department of Innovation, Industry, Science and Research (DIISR) in June 2010, *Investment Plan for the Research Capability AURIN*, in which the rationale for the AURIN project was set out and priorities for investment were canvassed, including the leveraging of other investments. While the research infrastructure being developed and operationalized by AURIN is focused on addressing issues of national significance, explicit attention is being given to demonstrate the application of the research infrastructure at various levels of scale and in varying scope across the nation's complex urban system. Importantly, the e-Infrastructure is incorporating not only historic and contemporary data, but also information based on forecasts and projections relating to future potential or possible scenarios about the development of the national urban system. This is particularly important in the context of researchers being able to use the research

infrastructure to inform the discussions occurring in Australia on the nation's future development and how that might reinforce or change current patterns and trends.

Objective

The objective of the AURIN project is to provide urban and built environment researchers with a portal providing seamless access to data and tools for interrogating a wide array of distributed data sets to support multiple research activities that will enhance the understanding of key issues across the urban components of Australia's settlement system. This includes providing an improved evidence base to inform policy and inform business decisions across urban settlements which is particularly important given national debates on issues such as the implications of national population growth; how that might be distributed across space; the types of urban environments in which diverse peoples might live, and how the nation might achieve the sustainable development of its cities and towns.

In that context e-Research (or e-Science) has much to offer urban and built environment researchers through software solutions addressing the heterogeneity of distributed data sets and associated metadata thus facilitating data discovery, supporting finer-grained access control and single sign-on reflecting the autonomy (security) of organizations involved, and inter-operability across them in a coordinated manner, while supporting accounting and auditing information on access to and usage of resources. Definition and enactment of workflows that allow researchers to share and repeat the way in which they currently access and use a wide variety of data sets and associated tools is also a highly desirable capability to support.

Existing Initiatives

In undertaking the AURIN project much can be learned from existing recent initiatives in e-Science enabling access to distributed data sets. That includes, for example, the ESRC funded *Data Management through e-Social Science* project (DAMES - www.dames.org.uk) (Tan et *al.*, 2009); the *National e-Infrastructure for Social Simulation* (NeISS – www.neiss.org.uk) (Birkin et *al.*, 2010), the *Transport Analysis Simulation System* (Transims)(Nagel et *al.*, 1999), and *Epidemic Simulation System* (EpiSims) tools (Barrett et *al.*, 2008), the European CESSDA efforts to harmonize social science archives across Europe (www.cessda.org), and the EU INSPIRE initiative to support global geospatial data initiatives (http://inspire.jrc.ec.europa.eu).

Limitations

EIF funds may not be used to fund research activity *per se*, nor for applications beyond those required for testing the AURIN e-Infrastructure in order to demonstrate its operability. However, it is recognised that some human resources are essential for the creation and development of research infrastructure and for some on-going activities, provided that they are considered critical to the creation and development of the infrastructure itself - for example, negotiating access protocols, design and software development activities, and data enhancement, design, coding and testing.

Furthermore, AURIN's funding can only be applied to the 'creation and development' of research infrastructure and may not be used for recurrent costs, such as on-going maintenance of infrastructure or operating costs. Therefore, the importance of co-investment is emphasised. On-going investment required to sustain and develop new AURIN research infrastructure and continue to provide services will need to be clearly identified as an input to forward planning. It is important to ensure the scope of the activity is realistic and achievable within the limited time-frame and with the funding available and the co-investment that is leveraged. Thus, incorporating and building on existing initiatives and capabilities will be important.

DEVELOPING THE AURIN WORK PROGRAM

The development of the AURIN work program from September 2010 to April 2011 involved extensive consultation with potential urban and built environment stakeholders through a process that included a series of roadshows, forums teleconferences and written submissions which followed the initial release of the *AURIN Investment Plan* in 2010.

In early May 2011 the funding agency (DIISR) approved the *AURIN EIF Final Project Plan* which, *inter alia*, set out the details of the work program for the project, with two fundamental deliverables for the AURIN project:

- The **Technical Architecture** to which approximately 35% of AURIN EIF funds are allocated; and
- The **Strategic Implementation Streams**, called **Lenses** which relate to issues of national significance, and to which approximately 55% of AURIN EIF funds will be allocated.

The AURIN Technical Architecture focuses on providing a single portal for researchers to access diverse data from a wide range of data providers, with the capability to integrate data and interrogate data through a 'toolkit' of analytical, modelling and visualization tools, thus providing access to a range of e-research capabilities.

The Lenses on which the AURIN research infrastructure is to relate include:

1. *Population and demographic futures and benchmarked social indicators.*
2. *Economic activity and urban labour markets.*
3. *Urban health, well-being and quality of life.*
4. *Urban housing.*
5. *Urban transport.*
6. *Energy and water supply and consumption.*
7. *City logistics.*
8. *Urban vulnerability and risks.*
9. *Urban governance, policy and management.*
10. *Innovative urban design.*

These Lenses have been selected by the AURIN Management Board as an 'aspirational' list that will cater for a significant range of interests among Australia's diverse urban and built environment research community. Work on the Lenses is being undertaken in a sequenced order and is subject to review by the Management Board. This may result in an amendment to the number of Lenses to be undertaken in light of the experience derived from implementing the early Lenses.

HOW AURIN OPERATES

The AURIN project is conducted in accordance with the AURIN Funding Agreement between the Commonwealth Government and the University of Melbourne as the Lead Agent. A capability Office for AURIN has been established in the Faculty of Architecture, Building and Planning at the University of Melbourne. The governance of AURIN is through a Management Board, which has appointed a Technical Committee to provide guidance and advice.

In developing the work program for AURIN as reflected in the *AURIN EIF Final Project Plan*, a comprehensive consultation program was undertaken between September 2010 and March 2011 engaging stakeholders. Consultation with stakeholders will continue throughout the AURIN project. The consultation process proved most valuable in many ways, including the identification of key groups of researchers and key potential data providers who have capability and willingness to become formally engaged in the project.

The intention is for AURIN to operate in a networked way embracing researchers across institutions and engaging data providers through Collaboration Agreements and Collaboration Sub-contracts. The AURIN Technical Architecture is being developed and operated out of the University of Melbourne but with

components of the e-Infrastructure being developed elsewhere through a range of Collaboration Sub-contracts.

Expert Groups have been established to assist with determining the scope, focus and situational context for the work on the Lenses. This work is being undertaken through engaging with key data providers and through specific projects undertaken through a series of Collaboration Agreements with data providers and Collaboration Sub-contracts with research groups building utilizing clearly demonstrated capabilities and existing expertise, and deliberately building on established collaborations and existing outputs that may be readily integrated and enhances within the AURIN e-Infrastructure.

THE CHALLENGES

A multitude of challenges confront the AURIN project. Some of those are outlined below, and they need to be addressed and accommodated within the AURIN e-Infrastructure.

1. *Urban systems are complex*. They encompass natural and built environmental, socio-demographic and economic elements, exhibit a wide range of human behaviours by the decisions of individuals, households, and businesses, and they are impacted by multi-levels of government and characterised by diverse governance processes – all of which can have a hierarchical structure, ranging from the large metropolitan cities to regional cities and towns, to very small urban localities. They are far from being static, constantly changing and adapting to economic, social and environmental, and institutional forces that are both exogenous and endogenous to the system.

2. *The urban and built environment research community in Australia is broad and diverse and is lacking in coherence*. That is not a criticism, but the reality. That community encompasses researchers from many disciplines, ranging from economists, geographers and sociologists to environmental health people, to traffic engineers, planners, architects and urban designers. It involves researchers who use data based on *de jure* regions, point-located data, and flows/network data. It encompasses researchers who operate with aggregated spatial data to those who operate at the micro-level using cadastre-based data. It encompasses researchers who use complex multi-variate spatial econometric modelling tools to those who focus on volumetric modelling of the built environment. It encompasses those whose research is embedded in scientific modes of analysis using quantitative analysis, hypothesis testing, and the collection of data through probability sample survey approaches to those whose research is embedded in qualitative analysis using approaches such as collecting oral histories and using case studies. And importantly the research community is increasingly making use of advanced geographic information technologies.

3. *The situational focus of research in the urban and built environment field is also highly variable*. It ranges from a national or state-wide perspective (where the focus might be on the analysis of regional variations in phenomena such as social advantage/disadvantage) to a specific metropolitan regional perspective (where the focus might be on small area characteristics of phenomena such as house prices or travel patterns of members of households), to a specific area or location perspective (where the focus might be on a specific building site, or a redevelopment project site, or a neighbourhood).

4. That means that *diverse types of data at different levels of spatial scale will need to be acce*ssed. In general, much of the data at a spatial area level that is used by urban researchers is embedded in one or more levels of the official geography for which census data is available from the Census Collector District (CCD) level up through Statistical Local Areas (SLAs), Local Government Areas (LGAs), Statistical Divisions (SDs), etc. That geography is widely used by Commonwealth and by state and territory agencies for administrative and other data collections. However, significant components of the urban research community and especially the geomatics, planning the built environment research communities are highly dependent on accessing highly disaggregated spatially including at the cadastre level.

5. There is the issue of the *temporal nature of data* used by urban and built environment researchers. While cross-sectional data relating to a specific point in time suffices for some, much research is dependent on access to time-series data. There is also a need for some research to access data that is in the form of *future estimates/projections/ forecasts*, and those data are rare and not systematised in collection and availability.

6. The *integration of national survey-based data with spatial data* derived from the census using micro-simulation methods (as demonstrated by the work of NATSEM) is now enabling the generation of synthetic estimates at the small area level of the likely incidence of survey variables. There is also an increasing interest among some researchers to be able to integrate unit records administrative and/or survey-based individual data with spatial objective data to incorporate situational data variables into behavioural data sets to enhance the modelling of relationships between behaviour and the environmental context in which it takes place. But wider researcher access to such micro- unit-level data is typically restricted to preserve confidentiality. The challenge is to facilitate that through securitized access systems.

7. While Australia is relatively well served in the provision of aggregated level data, such as that available through the census, which is (now) available in the context of a data 'creative commons', much of the data that would enhance research capability and capacity in the urban and built environment research community is 'locked-up' or is far from being readily accessible which represents a substantial impediment - or even a barrier - for research innovation.

Thus, AURIN is confronted with the *heterogeneity of researcher needs* and with the *heterogeneity of data that needs to be accessed across a diversity of sou*rces from the public, commercial and academic domains. Those data are available or potentially available in a range of formats. It is fair to say that, till now, *ad hoc* data access and management solutions have been the primary way in which urban and built environment research has been undertaken. Organizations have created silos of data with their own heterogeneous access demands and data models that prohibit or severely restrict researcher access and use by the wider community. Researchers subsequently have been left with swathes of distributed data sets from multiple organizations. Assuming we are aware of their existence and that access might be gained, researchers are typically tasked with manual collection and manipulation of those data through their own in-house techniques and their own usually limited expertise using a range of statistical and other interrogation tools. Furthermore, the data can take many forms and temporal scales, from archived social science resources to short-lived (ephemeral) data sets captured in real time.

AURIN e-INFRASTRUCTURE

Requirements

The AURIN research infrastructure needs to meet a range of needs from the urban and built environment research community and institutions and organizations and individuals that are stakeholders. In brief, this involves:

1. *Single sign-on*: There needs to be a single sign-on through the AURIN portal for those seeking to access the distributed, heterogeneous data sets and e-research tools available through AURIN,that is, access to many different secure systems and organisations can be achieved without the need for multiple authentication and/or authorization steps.

2. *Autonomy*: Data providers need to be able to define and enforce their own local discretionary policies regarding data access and usage of their own resources. In some cases data providers may choose to provide open access, in others managed and/or restricted access will be required. The AURIN e-Infrastructure will manage the realization of those data provider requirements.

3. *Usability*: The heterogeneous nature of distributed resources and their associated (if any) security needs to be made transparent to the end user researcher by the AURIN e-Infrastructure facilitating

access for the user with minimal knowledge of the middleware and technical solutions used to deliver that access. All that is needed will be an internet browser.

4. *Accountability*: In some cases users may need to be aware of the obligations arising with access to and usage of certain data from particular suppliers, especially for access to and use of sensitive data and individual record data, and the consequences for misuse. Managing this requires securitized access protocols and procedures which the AURIN e-Infrastructure will explicitly provide. Capturing access and usage therefore will be an essential component of the e-Infrastructure.

5. *Openness and inter-operability*: Where possible the AURIN e-Infrastructure will use open source solutions exploiting and supporting open standards and programming interfaces leveraging best practice.

6. *Collaborative nature*: The AURIN e-Infrastructure needs to support multiple research collaborations that may evolve over time, with researchers able to contribute in a range of research initiatives through the AURIN e-Infrastructure and input their own thoughts and ideas to its evolution.

Many of the above may be addressed through the development and support of e-Science virtual organizations (VOs) - in the AURIN context the Lenses - and especially those based around data-driven collaborations. For AURIN it is important those be dynamic and evolving, not simply offer/support static environments giving access to hard-coded data sets and resources. Researchers should be able to use the VO resources as information to access data as well as feed the VO with data sets themselves with results and comments on data they have interrogated, analyzed and visualized through the AURIN e-Infrastructure.

Design

Figure 1 provides a simplified view of the design of the technical architecture for AURIN. For the purpose of illustration, in the figure the focus is on energy and water (EW) research as an example, but the discussion is generic and could be used in and research scenario.

The AURIN e-Infrastructure is currently accessible through a LifeRay-based portal (www.liferay.com) which provides a platform and framework for collaborative and social application and services. It provides wiki and discussion forum support, the latest portal specification, and offers built-in support to a number of core user management systems. The AURIN portal has been deployed within, and is made available, through the Australian Access Federation (www.aaf.edu.au) to support federated authentication.

The e-Infrastructure architecture adopts a typical *n*-tier service-oriented architecture encompassing both traditional Simple Object Access Protocol (SOAP)-based web services and Representational State Transfer (REST) based web services accessible over *http*.

Data sets are accessed, used and manipulated through targeted JSR-286 compliant portlets that interact with a range of distributed services (S_{EW}) developed and deployed as part of the e-Infrastructure. Many AURIN services will leverage hosting services offered through distributed sites (EW_{Prov}) that potentially have their own access and usage policies realized through policy enforcement points (PEP) and policy decision points (PDP) associated with the specific services offered. In realizing this we leverage a body of work in security authorization (Sinnott, et a.2008; Watt, et al. 2011; Watt and Sinnott 2011; Wei, et al.. 2011) and the exploitation of security attributes delivered by the AAF. In particular, we exploit the *auEduPersonSharedToken* delivered as a core attribute in the SAML exchange between remote Identity Providers and AURIN portal (Service Provider) that exists within the AAF. Collectively the privileges, roles, and manipulation of attributes for user identity within the portal and for access to remote resources are expected to utilize a range of VO_{MGT} tools (Alfieri, et al. 2004). It should also be noted that for those sites without an IdP, e.g. commercial or other non-academic research partners, a virtual home is offered; that is, an IdP specifically set up for AURIN that can be used to authenticate and authorize individuals

from organizations not involved in the AAF directly. This is offered as a Virtual Home Organisation as part of the AAF.

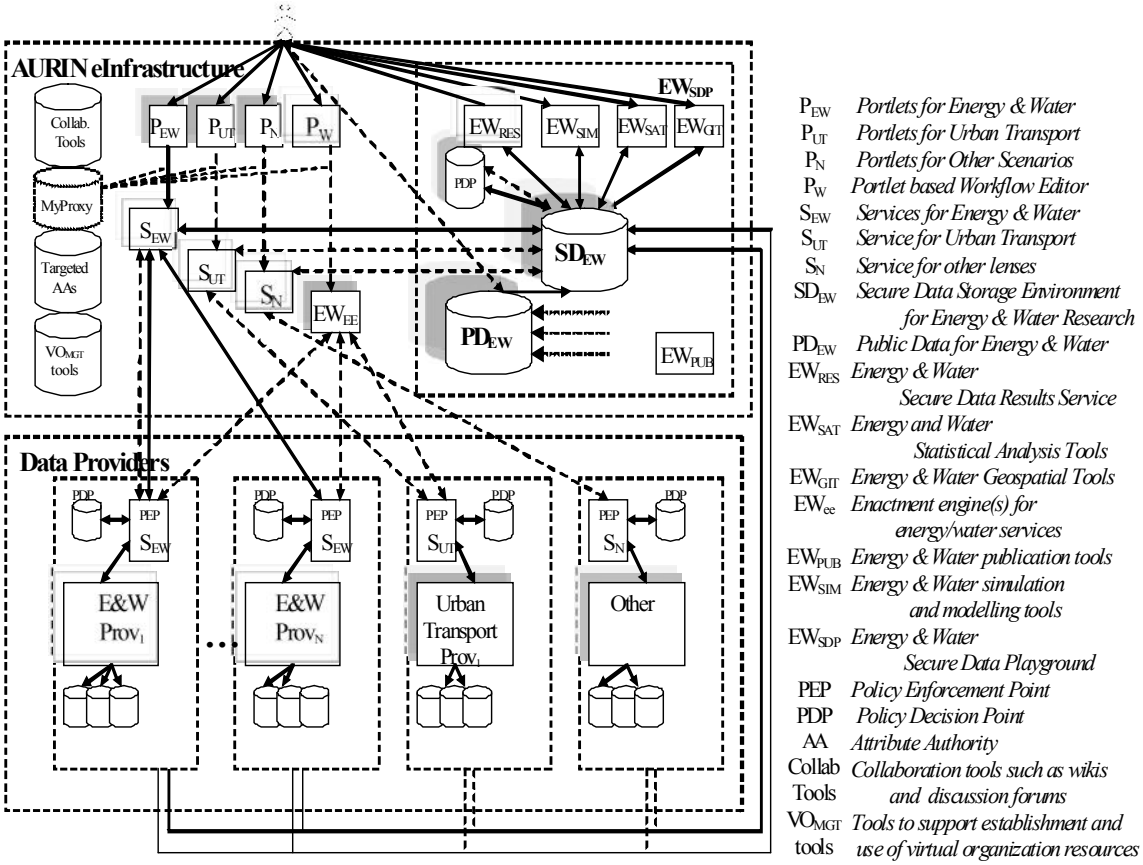

Figure 1: Simplified model of the AURIN e-Infrastructure

Data access and delivery in AURIN is based on agreed interfaces to known data sets and services realizes through pre-agreed interfaces (portlets). A variety of such services for energy and water and providers (S_{EW}), urban transport data providers (S_{UT}), or any other data/service providers (S_N) will exist depending on the nature of the research question/implementation streams for AURIN.

Thus, when a user logs-in to the AURIN portal through the AAF they are presented with different potential views of data sets made available by providers. This is a view of variables that can be selected, not the variables themselves. For security-oriented scenarios, depending on the nature and source of the data, it might be the case that the data sets are not immediately accessible, and might only be accessed and analyzed through a pre-defined collection of tools offering e-Infrastructure specific (fixed interfaces) through which data processing and analysis can take place.

By way of an example, for a particular energy and water scenario this might be a predefined library of statistical analysis functions in R, Stata, SPSS, etc., that are offered through a portlet interface that allow a pre-determined set of coding and analysis capabilities related to the energy and water scenario; for example, recoding routines to take local geospatial data sets (CCD, other boundary data sets, etc) from results data sets and map those onto larger LGA coordinates, or for recoding specific variables according to, for example, particular health classifications.

Data playgrounds are a key enabler for this kind of scenario to accommodate data providers wary of 'giving their data away'. Such data playgrounds can also act as a source of public data sets (PD_{EW}); for example, by mirroring of associated data from known energy and water data providers or uploaded by members of the research community. Data playgrounds can be realized in several different ways depending upon the nature of the data itself. An example might be a secure file system accessible to those energy and water researchers and data providers with sufficient access and usage privileges, or a DBMS reserved for researchers in this domain. The secure data playground will, where appropriate (and subject to security constraints for example) support larger scale data processing capabilities; for example, building indexes across data sets for faster searching. Extraction of metadata from these data sets will be a key aspect of the AURIN data management model and used for discovery, access and analysis of urban and built environment data sets.

As illustrated in Figure 1, in the AURIN architecture, a particular secure data playground (EW_{SDP}) is established where the data sets are aggregated by the service (S_{EW}) in some manner. This aggregation depends upon the data sets and the context of their usage in the AURIN e-Infrastructure. Where the data sets have a common index - for example, they can be linked and joined, this linkage information can subsequently be removed to avoid data disclosure risks. It is quite possible to have more than one secure data playground in the same VO where different users will have access to and use of their own individualized data sets. Each of these data playgrounds will have their own security constraints and access and usage policies.

It has been recognized that the collection of services comprising the AURIN e-Infrastructure needs to evolve over the course of the project and that researcher's wish to coordinate the way in which these services are accessed and used in an orchestrated and community-driven manner. To this end, support for workflows that allow coupling of the interactions between collections of services and data movement between services and the secure data playground has been identified as beneficial. There are many workflow environments that currently exist and indeed portal-based solutions that allow for workflows to be enacted (Kacsuka, et al., 2008). Support for targeted enactment engines (EW_{EE}) that allow, for example, AURIN energy and water (or indeed other) workflows to be defined, shared and subsequently enacted by other researchers is a key goal of the AURIN e-Infrastructure. Work is currently on-going in evaluating these environments, and especially with regard to the way in which they support finer-grained security. A summary of capabilities for security-oriented workflows in the social sciences is described in Sinnott and Hussain (2009).

THE AURIN RESEARCH DATA LANDSCAPE

Potentially the AURIN research data landscape is vast extending across a wide array of Commonwealth and State and Territory agencies, as well as local governments, commercial organizations, and, of course, researcher-generated data sets.

Many of these data providers have data that is directly accessible on the web. Typically this is through web pages that offer files and data for download. In the case of the Australian Bureau of Statistics (ABS - www.abs.gov.au), for example, there are over 1 million web pages through which a huge array of data can be accessed from the Census (and previous versions of the Census) to an extensive array of surveys and reports related to the Australian population. Similar data deluge scenarios exist with many other government and commercial organizations.

At the heart of the AURIN work is providing programmatic access to these data sets in a manner that supports the researchers and research process. It is recognised that simply giving access to all of the existing data (should this be possible!) is not the best way to add value to urban research since there is simply too much data and domain expertise required to understand and utilize the data effectively. Instead the project has identified a key set of strategic research areas to be realized through implementation stream (the 10 Lenses listed earlier) of importance to the urban research community. Each of these has their own data sets and services that need to be brought together. Each of these

lenses will be supported through a common underpinning e-Infrastructure, in effect providing a VO offering data sets, tools and security policies relevant to researchers and data providers in that domain.

Each of these areas represents a significant urban research area in its own right. However a key challenge (and research opportunity) is that all of these Lenses are themselves interrelated. As one example, understanding the changing profile of population demographics in cities and the current and future urban landscape is essential for planning urban transport, housing, energy and water, and provisioning of healthcare.

Key data providers

There are key data providers among public agencies with which AURIN needs work to broker data access collaboration agreements, and some of there have been identified as priority in the *AURIN EIF Final Project Plan* (2011). These include, *inter alia*, organisations such as: the Australian Bureau of Statistics (ABS); Geoscience Australia; the Australian Institute of Health and Welfare (AIHW); the Bureau of Infrastructure, Transport and Regional Economics (BITRE), the Population Health Information Development Unit at the University of Adelaide; and Public Sector Mapping Agency (PSMA). Those organisations can provide access to data at the full national level and at spatially disaggregated levels of scale.

As far as possible, AURIN also needs to broker data access collaboration agreements with key agencies in the States and Territories that deal with land data, planning, transport, water and energy, and health. In addition, AURIN needs to facilitate tapping researcher-generated data sets, particularly those derived through ARC and other competitive grant-funded research efforts. In some of the Lenses it will be necessary to work with private sector organizations, including through industry associations, to broker data access collaboration agreements. Thus the heterogeneity of the distributed data access challenge confronting AURIN is substantial.

Data diversity

Data that might be accessed through AURIN will be diverse:

1. Mostly it will be spatial data at various levels of scale, ranging across the vast volume of data that is embedded in the various levels of the census geography (CCDs, SLAs, LGAs, Postcodes, etc). Those data – sometimes referred to as macro data - are aggregated data and comprise the basis on which much urban research is conducted. Access to such data is relatively open – in what is a creative data commons environment. Such data might incorporate new functional geographies derived by researchers such as Functional Economic Regions (Liao, et al. 2009) and voting booth catchments (Stimson, et al. 2011).

2. Spatial data takes many forms; point located data; polygon (area) data; flows/network data; surface/trend surface data; etc.

3. Built urban and built environment research also requires micro data embedded in the cadastre. Such data is not particularly easily accessed, and much of it is locked-up.

4. Individual-level data that is unit record data collected through surveys and assembled by public agencies and service providers as administrative data bases that is highly relevant to urban and built environment research is generally not readily accessible. GIS technology and micro-simulation tools readily permit the integration of such unit record data with spatial objective data, but securitized access systems are needed to achieve that.

5. There may also be qualitative data that might be accessed through AURIN.

e-Research tools

Crucial to the utility of the AURIN e-Infrastructure will be the capability for data integration and interrogation through the provision of open source e-Research tools to facilitate:

- spatial analysis and modelling;
- statistical analysis and modelling;
- simulation modelling;
- volumetric modelling;
- map and graphic visualization; and
- generation of reports.

Open source e-Research tools need to be available as part of the AURIN e-Infrastructure to facilitate these data interrogations. Existing capabilities might be incorporated and enhanced within AURIN, such as the initiative by the former ARC Research Network in Spatial Social Science (ARCRNSISS) e-Research Facility for Socio-Spatial Analysis at the University of Queensland which enables data embedded in a level of the census geographies to be analyzed, modelled and visualized using tools available in an open source environment, such as multiple-regression, principal components analysis, multi-discriminant analysis, and cluster analysis, and using a range of visualization approaches from a menu of spatial categorization procedures.

THE LENSES

The 10 'aspirational' Lenses on which it is intended the AURIN research infrastructure will focus as issues of national significance help define key data types and research methodology approaches. Focusing effort within the topics identified by the Lenses helps identify the detail of the data systems and the analytical, modelling and visualization tools that will need to be incorporated within the AURIN e-Infrastructure. There will be consistent and standardized platforms compatible across the Lenses, as well as specific requirements that arise in response to issues specific to a Lens. The AURIN e-Infrastructure will be designed to allow flexibility for adaptation to new forms of data and to accommodate new needs anticipated within the Lenses.

It is important to note that the implementation of the Lenses is being driven by pragmatic considerations, including resource limitations, difficulties of accessing the required data, and the need to build on existing initiatives, capabilities, expertise and experiences. Furthermore, there will need to be an explicit selective focus within each Lens as so as to use the Lens as a demonstration project (or series of projects) developed to the stage where there is explicit testing to demonstrate the potential for the AURIN initiative to 'value add' to urban and built environment research capability. It needs to be recognized that the AURIN project will *not* be able solve all problems, *nor* meet the needs of every possible research interest with respect to a particular Lens.

Of course many data sets and some of the tools and protocols developed for value adding in the AURIN research e-Infrastructure will be common across multiple Lenses, and it is important that that be so.

Thus, there will be a disciplined prioritization of focus and activity for the implementation of each Lens. The Expert Groups established to provide guidance and advice on the design and implementation of each Lens will affect the focus of the activity. This will tend to vary in terms of scale, coverage and complexity, and in accordance with pragmatic considerations concerning the difficulties of accessing the required data.

For example, it may be possible for the data sets to be incorporated into the e-Infrastructure for a Lens to be fully national in coverage at, say the level of scale of Statistical Local Areas (SLAs). Alternatively, for demonstration purposes it may be decided to focus attention on the operationalization of a Lens to be limited to one or two metropolitan cities because of co-investment from participating agencies providing data, and in order to demonstrate the application of specific data interrogation tools to enable important hypotheses to be tested or for simulations to be run through the enabling e-Research capabilities

provided in the AURIN e-Infrastructure. In that way, for each Lens there is a process of prioritization for the allocation of the investment informed through a Lens management process involving advice from the Lens Expert Groups.

The implementation of a Lens will incorporate explicit demonstrators in order to show explicitly how the AURIN e-Infrastructure can operate through access to distributed data, data integration, and e-research tools to facilitate data interrogation through analysis, modeling and visualization to produce outputs that can be of value to inform policy and planning.

Collaboration Agreements and Sub-contracts

Collaborating organizations/institutions and co-investors are being engaged through Collaboration Data Access Agreements and Collaboration Sub-contracts to bring their specialist knowledge and experience, and their history of project delivery, to participation in particular Lenses or sub-components of the Lenses appropriate to their expertise. The sub-contracting process necessarily involves:

- directed invited proposals/tenders where there is an obvious data and/or service provider and/or an established expertise;
- calls for expressions of interest from stakeholders; and
- open competitive bids.

Those Collaboration Agreements and Collaboration Sub-contracts for components of the work program work relating to each Lens required demonstration, where appropriate, of:

- cross-institutional collaboration
- relevant collaboration with data providers and public and private partners where appropriate;
- a demonstrated track record relevant to delivery of outputs
- co-investment, which may take multiple forms (cash, in-kind, a valuation of data and tools, etc.).

Staged implementation

There is to be a staged implementation of the AURIN Lenses.

This commenced in June 2011 with the formation of the Expert groups for the first three lenses - Lens 1: *Population and demographic futures and benchmarked social indicators*: Lens 2 *Economic activity and urban labour markets*; and Lens 3: *Urban health, well-being and quality of life, and Social data and benchmarked indicators*. Work on those Lenses commenced in September/October 2011 and will progress over a period of 12-14 months.

The planning process begins in September/October 2011 for Lens 4: *Urban housing* and Lens 5: *Urban transport*, followed by Lens 6: *Urban Water and electricity supply and consumption*. Work on those lenses will be initiated from about February/March 2012 and will run for a period up to mid-2013.

Planning for Lens 7: *City logistics* and Lens 8: *Urban risk and vulnerability* will commence about April 2012 with work on those Lenses from about August 2012 and will continue to the end of 2013.

Planning for the final two Lenses, Lens 9: *Innovative urban design* and Lens 10: *Urban governance, policy and management*, will begin in February 2013, with implementation of work in June 2013 and completion in June 2014.

The initial 3 Lenses

Work on the first three Lenses has commenced within a framework illustrated in Figures 2, 3 and 4.

As shown in Figure 2, Lens 1: *Population and demographic futures and benchmarked social indicators*, explicitly has a national-coverage focus drawing on data available largely from key national data

agencies, including the ABS census, plus researcher generated data sets which have produced value-enhanced data, such as that derived through micro-simulation methods.

Access to such data is crucial for urban researchers to conduct analysis and modelling of phenomena such as migration flows and spatial differentiation in the incidence of territorial social indicators across a range of social, economic and other domains in order to better understand issues related to social inclusion/exclusion and social advantage/disadvantage. In addition, it is crucial that urban researchers have access to population and demographic forecasts/estimates in addition to the normal time-series and cross-sectional data. Those diverse data need to be available at a variety of spatial scales for interrogation by researchers in a national, state and territory, metropolitan city and other regional scale, and those data need to be benchmarked for both national and international comparative purposes. The AURIN e-Infrastructure needs to facilitate researcher interrogation of such data to conduct spatial and statistical analysis and modelling using a suite of e-research tools with advanced visualization of both data inputs and the outputs of analysis and outputs.

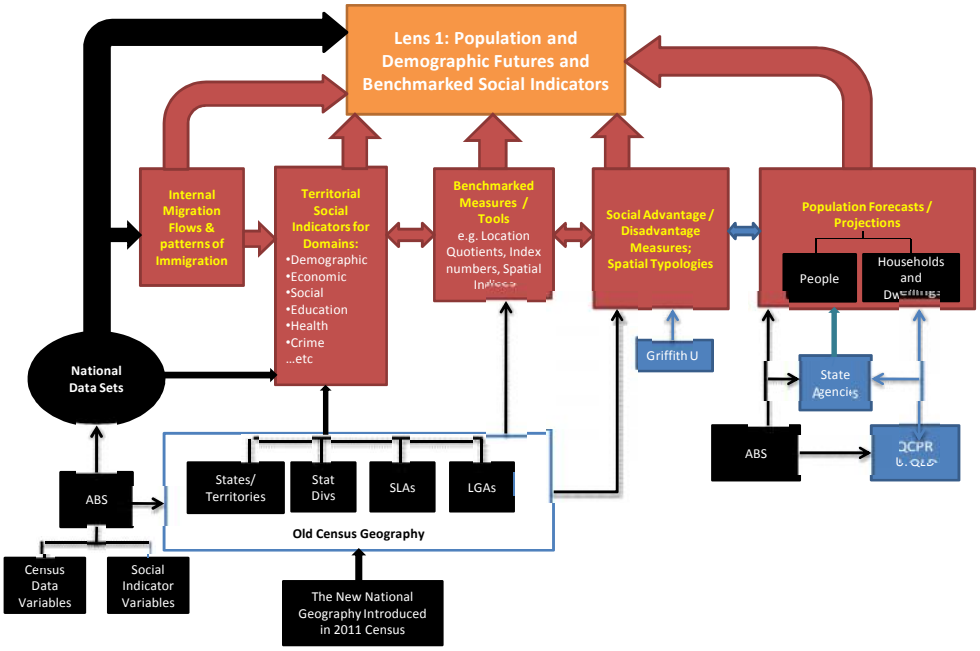

Figure 2: Framework for Lens 1

Lens 1 is being operationalized within the framework of the official national geographies relating to the census.

It is self-evident that the data accessible and available for interrogation through Lens 1 will be of considerable values as data inputs for many of the other AURIN Lenses, such as using small area population projection data to help model future demand for household consumption of water and electricity in a metropolitan region.

As shown in Figure 3, Lens 2: *Economic activity and urban labour markets*, has both a national coverage focus and an explicit focus on metropolitan regions, drawing on data that needs to be sourced from a number of national and state-metropolitan level data sets as well as researcher generated data sets. In addition, commercial data sets will be relevant.

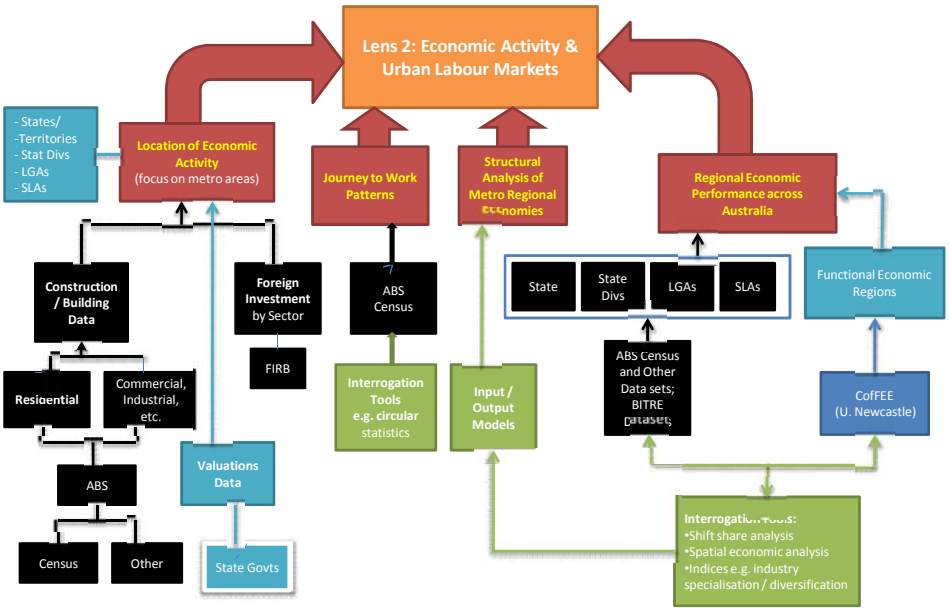

Figure 3: Framework for Lens 2

Urban research in this Lens focuses on the analysis and modelling of data that is embedded within the official census geographies; but it also utilizes researcher-generated geographies such as Functional Economic Regions (Stimson, et al. 2011). Those data need to be interrogated using suites of e-Research tools permitting researchers to conduct spatial econometric analysis and modelling of aspects of regional economic performance and investigate the relationships between where people live and work using journey-to-work matrices.

Additionally, spatially disaggregated data at the cadastral level in addition to spatially disaggregated data at the small area level in the census geographies might be integrated through Lens 2 to permit researchers to investigate the spatial patterning of the location of economic establishments in particular industries within a city in the context of land use planning.

The outputs derived from interrogating data accessed through Lens 2 might be used as inputs for research using data in other Lenses, such as investigating the relationship between labour markets and housing markets.

As shown in *Figure 4, Lens 3: Urban health, well-being and quality of life*, is being operationalized in both a national coverage context and a situational-specific context.

Thus on the one hand this Lens is providing researchers with access to spatially disaggregate data that provides information on population health related variables derived from national data, whilst on the other hand, the Lens is providing researchers with cadastral-based data specific areas that enables individual health data to be linked to environmental data thus allowing researchers to generate information such as a walkability index for specific geographical location. This will subsequently allow researchers to model the relationship between, say, walking behaviour and local urban morphology characteristics and health outcomes. The investigation of such relationships which the AURIN e-Infrastructure may facilitate are being achieved through specific demonstrator projects for this Lens.

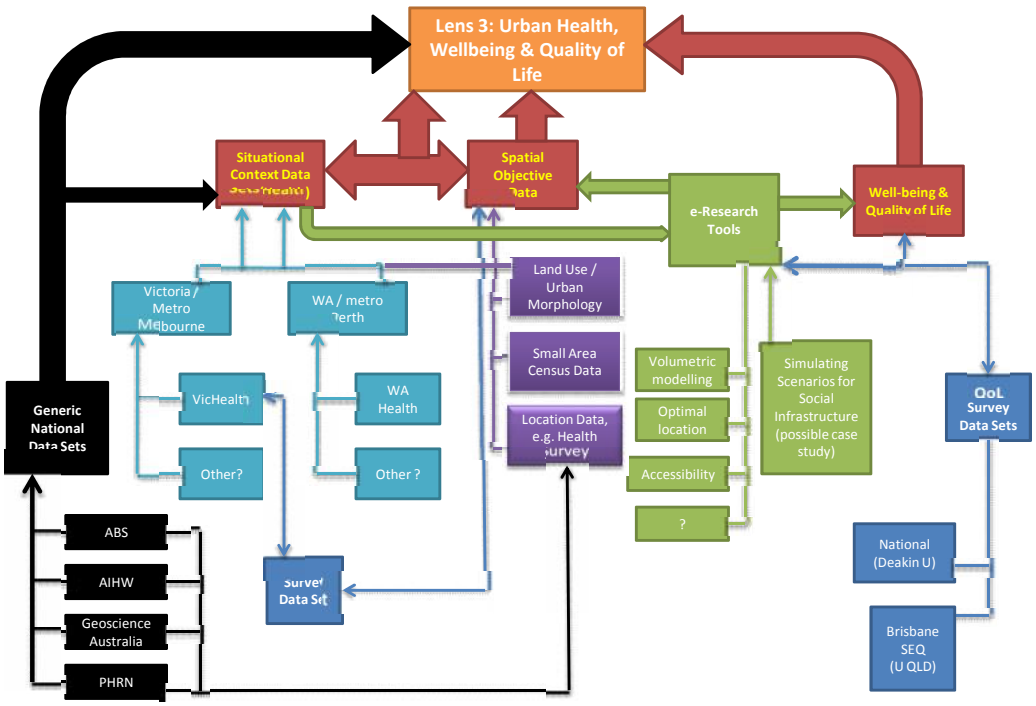

Figure 4: Framework for Lens 3

In addition, Lens 3 incorporates a demonstration of how individual record data collected through a large-scale household survey where the residential location of survey respondents is geocoded may be integrated with non-survey spatial objective data to add variables to the survey data and also permit, through the use of GIS tools and survey data, spatially-generalized representation of behavioural phenomena.

Finally, the Lens will demonstrate how survey data relating to measuring of quality of life may be integrated with census data through micro-simulation methods to derive spatial area estimates of survey data phenomena.

EXPECTED BENEFITS

Benefits flowing from the AURIN project are expected to include the following:

1. An enhanced capability to understand the physical and social aspects of built environments, including urban resource use and management, to improve the sustainability of cities and urban areas.

2. Portal-based access to an information infrastructure enabling the future development of research to underpin urban resource management and policy decision-making that will support greater sustainability (environmental, social and economic) of Australian cities and the urban components of its non-metropolitan regions.

3. An enabled (improved) collection and integration of disparate urban and built environment research datasets and support for researchers accessing and interrogating them including using e-Research capabilities including statistical and spatial analysis and modelling tools and visualisation tools.

4. An improved ability to locate, analyse and disseminate built environment and urban research datasets, including the development of protocols for integrating, analysing and accessing data.

5. An enhanced capability to undertake collaborative research domestically and on an international level.

Ultimately, a key barometer to judge the success of the AURIN project will be how researchers themselves move from the current *de facto* individual user-based approaches when undertaking research to an on-line collaborative research approach built upon a common, integrative research environment.

REFERENCES

Alfieri,R., Cecchini, R., Ciaschini, V. and Dell'Agnello, L. (2004),VOMS, an authorization system for virtual organizations, LNCS 2970, pp. 33-40.

AURIN EIF Final Project Plan (2011), April (available on www.aurin.org.au)

Barrett, C. L., Bisset, K. R., Eubank, S. G., Feng, X., Marathe, M.V. (2008) *EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks*, Proceedings of the 2008 ACM/IEEE conference on Supercomputing, ISBN: 978-1-4244-2835-9.

Birkin, M., Allan, R., Beckhofer, S., Buchan, I., Finch, J., Goble, C., Hudson-Smith, A., Lambert, P., Procter, R., de Roure, D., Sinnott, R.O. (2010) The Elements of a Computational Infrastructure for Social Simulation, Journal of the Philosophical Transactions of the Royal Society A, July 2010, (DOI:10.1098/rsta.2010.0150).

Kacsuka, P., Kiss, T. and Sipos, G. (2008),*Solving the Grid Interoperability Problem by P-GRADE Portal at Workflow level*, Future Generation Computer Systems, 24 (7).

Tan, L., Lambert, P., Turner, K. J., Blum, J., Bowes, A., Bell, D., Gayle, V., Jones, S. B., Maxwell, M., Sinnott, R.O., Warner, G. (2009) Enabling Quantitative Data Analysis through e-Infrastructures, Social Science Computer Review, January.

Liao, E., Shyy, T-K. and Stimson, R.J. (2009) Developing a web-based e-research facility for socio-spatial analysis to investigate relationships between voting patterns and local population characteristics, *Journal of Spatial Science*, 54(2) pp. 62-88.

Nagel, K., Beckman, R. J., Barrett, C.L. (1999) *TRANSIMS for urban planning*, 6th International Conference on Computers in Urban Planning and Urban Management, Venice, Italy, September 1999.

Sinnott, R.O., Chadwick, D. Doherty, T., Martin, D., Stell, A., Stewart, G., Su, L. and Watt, J. (2008), Advanced Security for Virtual Organizations: Exploring the Pros and Cons of Centralized vs Decentralized Security Models, *8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2008),* May, Lyon, France.

Sinnott, R.O. and Hussain, S. (2009), Architectural Design Patterns for Security-oriented Workflows in the Social Science Domain, *Proceedings of International Conference on e-Social Science*, Cologne, Germany, June.

Stimson, R., Mitchell, W., Rohde, D. and Shyy, T-K. (2011), Using Functional Economic Regions to model endogenous regional growth: implications for addressing the spatial autocorrelation problem, *Regional Science Policy and Practice*, 3 (3), pp. 131-144.

Watt, J. and Sinnott, R.O (2011), Supporting Federated Multi-Authority Security Models, *Proceedings of the 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, Newport Beach, CA, USA. May.

Watt, J., Sinnott, R.O., Inman, G. and Chadwick, D,W. (2011), Federated Authentication and Authorisation in the Social Science Domain, *Proceedings of the Sixth International Conference on Availability, Reliability and Security*, Vienna, Austria. August.

Wei, J. Arshad, J. and Sinnott, R.O. (2011), A Review of Grid Authentication and Authorization Technologies and Support for Federated Access Control, *Journal ACM Computing Surveys*, 43 (2), January.