The Urban Research Gateway for Australia:
Development of a Federated, Multi-disciplinary Research e-Infrastructure

Richard .O. Sinnott[1], Christopher Bayliss[1], Andrew Bromage[1], Gerson Galang[1], Guido Grazioli[1], Philip Greenwood[1], Angus Macauley[1], Damien Mannix[1], Luca Morandini[1], Marcos Nino-Ruiz[1], Christopher Pettit[2], Martin Tomko[1], Muhammad Sarwar[1], Robert Stimson[2], William Voorsluys[1], Ivo Widjaja[1]
[1]Department of Computing and Information Systems
[2]Faculty of Architecture Building and Planning
University of Melbourne, 3010 Victoria, Australia
Contact Author: rsinnott@unimelb.edu.au

Abstract—**The $20m Australian Urban Research Infrastructure Network (AURIN) project (www.aurin.org.au) began in July 2010. AURIN is developing a secure, web-based virtual environment (e-Infrastructure) - a *lab-in-a-browser* - offering access to diverse, distributed and extremely heterogeneous data sets together with an extensive portfolio of targeted analytical and visualization tools. This is being provisioned for Australia-wide urban and built environment researchers – itself a highly heterogeneous collection of research communities with diverse demands. This paper describes these demands and their associated needs and expectations on the e-Infrastructure and illustrates through a range of working examples how the e-Infrastructure allows inter-disciplinary research collaborations to take place. An overview of the e-Infrastructure itself is provided and how it allows tackling these demands.**

Keywords: Urban Research, e-Social Science, e-Health, e-Planning

I. INTRODUCTION

The Australian Urban Research Infrastructure Network (AURIN) project (www.aurin.org.au) is a major national project across Australia that commenced formally in July 2010. AURIN received $20 million of funding from the Australian Government Department of Innovation, Industry, Science Research and Tertiary Education (DIISRTE – www.innovation.gov.au) for the '*establishment of facilities to enhance the understanding of urban resource use and management*'. In particular, the AURIN project has been tasked with providing urban and built environment researchers with a state of the art research infrastructure – an e-Infrastructure - offering seamless and secure access to data and tools for interrogating a wide array of distributed data sets from diverse agencies, to support a portfolio of research activities reflecting the diversity of the urban and built environment research agenda.

Australia, as indeed is the case with many other countries, faces numerous challenges in the growth and planning of its cities, yet there is surprisingly little integrated infrastructure that allow for the complex information that might inform policies and research agendas more generally to be accessed and processed for informed decision making based upon qualitative data. Instead a variety of largely *ad hoc* and non-interoperable infrastructures and data sets has been developed over time by a range of national and State-based governments (Victoria, etc), and indeed by commercial and research organisations. AURIN is tasked with breaking down the data and organisational silos that have grown over time and are largely a barrier to many eResearch endeavours. To improve the way urban research itself is conducted, it is essential to make accessible the silos of data that exist across Australia to overcome the internet-hopping *modus operandi* of research where researchers access a multitude of web based resources on a one-by-one basis, or often spend weeks/months in obtaining permission to access particular resources hidden behind organisational firewalls. To achieve this it is necessary to develop and support services that allow data discovery and *federated* data access, i.e. *in situ* access to data from the data providers. This federated model is essential for many reasons. For many data sets, e.g. individual unit records or data from commercial organisations, it is simply not tenable to build a centralised data warehouse for all urban data. Furthermore as data grows and evolves over time it is highly beneficial to seamlessly leverage these updates and enhancements. Federated data access data models provide such opportunities that a centralised data warehouse does not.

The implementation of the AURIN e-Infrastructure commenced mid-2011, with the first year year of the project focused largely on gathering community-wide research requirements on the core capabilities and data sets that should be provisioned (made accessible) through the e-Infrastructure to the urban and built environment research community [1].

The University of Melbourne is the lead agent responsible for the successful delivery of the AURIN e-Infrastructure, however it is emphasised that the project is to be (is being!) developed and delivered in a networked manner – working with a multitude of agencies and groups across Australia providing either data or tools that should be integrated into the AURIN e-Infrastructure. The Melbourne eResearch Group at the University of Melbourne are primarily tasked with this integration effort.

The cornerstone of the AURIN e-Infrastructure is on providing programmatic access to a wide and heterogeneous array of data in a manner that supports urban and built environment researchers, as well as reflecting the agencies (government, commercial and academic) and associated stakeholders that are involved and especially their associated systems and processes. Thus AURIN cannot mandate that

complex AURIN-specific software systems/software stacks are installed and configured on government/commercial enterprise resources. Rather the AURIN e-Infrastructure has to be cognisant of the existing solutions already deployed by the organisations involved.

The field or urban and built environment research itself is very broad and covers a huge array of disciplines: population demographics, labour markets, socio-economics, health, transport, housing, amongst many other research dimensions. Specialisations of these are also commonplace. For example, a focus on indigenous populations, on the mental health of individuals living in cities, housing challenges facing first home buyers etc. To accommodate the challenge of developing an e-Infrastructure accommodating such diversity of research need, AURIN has identified a set of strategic implementation streams (*lenses*) of importance to subsets of the urban and built environment research community. Each of these lenses has their own data sets, services and tools that need to be provisioned. The set of AURIN lenses that were originally identified in the AURIN business plan included:

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.

However driven by guidance by the AURIN management board who provide oversight and independent guidance on the AURIN project as a whole, the lenses associated with city logistics, urban vulnerability and risks, and urban governance, policy and management have been removed from the current phase of the work. This was in part due to the complexities in gaining access to the necessary data as well as the significant amount of on-going sub-projects associated with AURIN across the existing lenses. It is anticipated that over 50 separate subprojects will be sponsored through AURIN, that the Melbourne eResearch Group are tasked with integrating into a unified e-Infrastructure.

The purpose of this paper is primarily to illustrate the application of the AURIN e-Infrastructure as a unified scientific gateway for urban research across Australia highlighting the diversity of the data and tools that are currently available and their usage across a range of urban research endeavours. Detailed information on the original proof of concept AURIN implementation was described in [2], with the extended data-driven AURIN solution described in [3,4]. The security solutions that are being rolled out across AURIN are described in more detail in [5]. The detailed enumeration of the AURIN project portfolio that is to be integrated into the AURIN platform is discussed in [6]. The use of Cloud resources and performance measurements of using such facilities for enacting urban workflows is described in [7-9].

The rest of the paper is structured as follows. Section 2 provides a summary of the core features of the technical architecture. Section 3 provides a summary of the Australian data landscape. Section 4 illustrates through a series of examples, how the AURIN e-Infrastructure can be utilized to support urban research endeavours. Section 5 focuses on related work undertaken in the urban research space and draws some conclusions on the work as a whole highlighting areas of future work.

II. AURIN E-INFRASTRUCTURE

The vision of the AURIN e-Infrastructure is to provide a unified environment for urban and built environment research. Whilst it is quite possible to develop a collection of heterogeneous collection of data services and resources targeted to subsets of the urban research landscape, AURIN was tasked with a grander vision: a unified and integrated environment that could be used for a multitude of urban research endeavours through a single one-stop-shop: the Australian urban research gateway as shown in Figure 1.

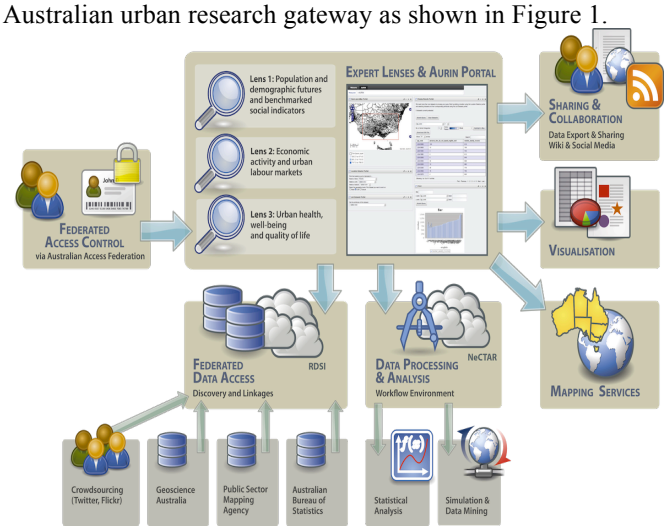

Figure 1: AURIN Architectural Vision

As presented in [3], the AURIN e-Infrastructure is being designed around a loosely coupled, flexible and importantly, an extensible service-oriented architecture-based paradigm. This extensibility is essential since the project continues to be tasked with providing access to and integrating a variety of new flavours of data beyond the traditional two-dimensional relational and structured data, as well as new services and tools.

To achieve this, the AURIN architecture is comprised of a range of components that communicate predominantly through Representational State Transfer (REST) based service calls. These calls leverage the JavaScript Object Notation (JSON) for their message format encoding through its support for hybrid messages with adaptive content. This is particularly advantageous for the complex data descriptions and formats to be passed around within the AURIN e-Infrastructure. In particular, given the natural geospatial application domain of

AURIN, the GeoJSON (www.geojson.org) data format has been used extensively for internal spatial data transfers between core architectural components.

The AURIN data e-Infrastructure extends the basic ideas of data Grid pioneered in earlier e-Science/eResearch projects such as [10-12] and is completely data driven. The access to and usage of data from heterogeneous data providers is driven by metadata that is automatically harvested from a rich variety of data service endpoints. Data can come in many flavours: structured data as might be found in a relational database through to unstructured data formats and 3D volumetric data. At the heart of the AURIN data-driven e-Infrastructure is a data registration service. This is accessible through a REST-based interface, exposing methods to read, write, modify and delete records (depending on user/data provider credentials). Registration of new datasets in the data registration database predominantly occurs through automatically harvesting and moderating the metadata from remote metadata service catalogues. A manual process is also offered. This includes support for bulk upload of data sets and importantly descriptions of their associated metadata. At present it is possible to harvest information from a portfolio of service endpoints including geospatial endpoints, e.g. Open Geospatial Consortium compliant Web Feature Services through to web services and even JDBC endpoints. These results are stored in an extensible (schema-free) structure. Through utilization of the open-source indexing system Solr (http://lucene.apache.org/solr/) the metadata allows for searching over a range of terms and variables – driven by the available metadata (see left of Figure 2) with the metadata highlighted (see centre of Figure 2) for a data set from the Victorian Department of Health and the kinds of information/variables that are available (see right of Figure 2) – in this case survey data on inadequate sleep is highlighted.

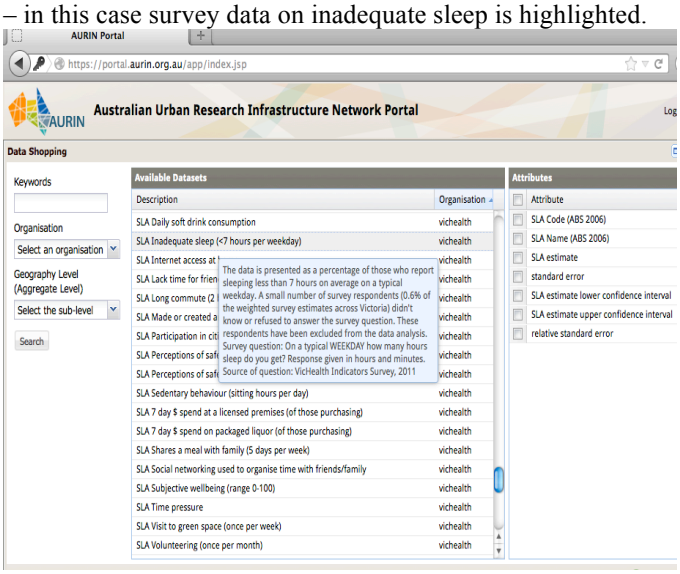

Figure 2: Data Request Interface and Use of Metadata

Urban research data is implicitly geospatial in its nature. Tools that allow filtering of the data based upon geospatial information / context are key to control the data deluge facing urban researchers [13]. However it is the case that a rich spectrum of geospatial information exists in many data resources and at a variety of scales: from latitude/longitudes, addresses, postcodes, Census districts, statistical local authorities (SLA), local government areas (LGA), cities, States, through to research defined geospatial areas such as labour force regions (LFR) and functional economic regions (FER). Other flavours of data also exist and must be managed by AURIN including social media data such as Twitter, graph based data, e.g. road networks, through to 3D data models of cities.

To tackle this the AURIN platform supports the filtering and selection of data sets based upon a range of geospatial aggregation levels and their subsetting as shown in Figure 3 where the selection of areas (and hence data of interest) is done at the LGA level for Victoria. The selection of areas of interest can be done through the user interface in several ways: through the pull down menus and selection of areas/geospatial data levels of interest, or through the map based interface highlighted in Figure 3.

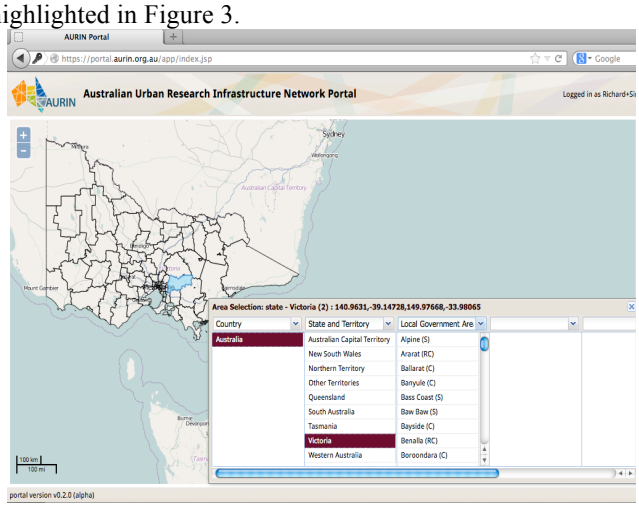

Figure 3: Selecting geospatial region and hence data of interest

Other core capabilities offered through the AURIN e-Infrastructure include:

- persistent data storage (storing GeoJSON formatted data objects);
- access to distributed data sets from a range of providers through an extensible array of data clients;
- geospatial services that provide capabilities to deal with the different geographic reference systems currently in use;
- access through the Australian Access Federation (AAF – www.aaf.edu.au) with work on-going to extend the basic authentication model of the AAF to incorporate more advanced authorization capabilities [5];
- an advanced user interface including support for brushing and visualization;
- a range of analytical and visualization tools, and
- workflows utilizing the Object Modelling System version 3 (www.javaforge.com/project/oms) and described in more detail in [7-9].

III. AUSTRALIAN URBAN DATA LANDSCAPE

Urban research can be classified as data intensive research. Unlike other research disciplines where access to large-scale compute facilities is the primary hindrance to research breakthroughs or to enhance research efforts, urban and built environment research is stifled through both access to and understanding of data. As noted, across Australia a huge array of organizations exists that hold data that is fundamental to supporting urban research. Whilst many of these data providers often have data that is directly accessible on the web, e.g. the Australian Bureau of Statistics (ABS – www.abs.gov.au) has data for direct download from its website – typically as Excel spreadsheets or zipped files, this model of data delivery places major challenges for researchers when dealing with the volume and diversity of such data. As one example, the ABS has literally thousands of spreadsheets and .zip files available for download covering a wide spectrum of urban phenomenon. This situation is magnified when juxtaposed with other national and State-wide organisations holding data that can/should be used to influence urban research: Geoscience Australia (www.ga.gov.au); the Public Health Information Development Unit (PHIDU - www.publichealth.gov.au); the Bureau of Infrastructure, Transport and Regional Economics (www.bitre.gov.au); the Australian Institute for Health and Welfare (www.aihw.gov.au); the Australian Housing and Urban Research Institute (www.ahuri.edu.au); the Department Climate Change & Energy Efficiency (www.climatechange.gov.au); the Department of Sustainability, Environment, Water, Population and Communities (www.environment.gov.au) amongst others. At a State-based level other agencies hold a rich variety of data that can/should inform urban research: these include transport agencies (VicRoads - www.vicroads.vic.gov.au), health agencies (VicHealth - www.vichealth.vic.gov.au) and the Health department of Western Australia (WAHealth - www.health.wa.gov.au) amongst many others.

A further dimension to this data spectrum is that a multitude of commercial organizations also hold data sets that need to be unlocked for urban researchers, e.g. the Public Sector Mapping Agency (PSMA – www.psma.com.au) hold the definitive geospatial information for Australia; commercial utility companies such as Ergon (www.ergon.com.au) hold energy and water information whilst real estate companies such as the Australian Property Monitors (APM – www.apm.com.au) hold vast holdings of housing and rental data across Australia.

Overcoming this diversity is at the heart of the AURIN e-Infrastructure. Urban researchers should be able to access diverse data sets as simply as possible. Key to this is the notion of single sign-on where users authenticate through the AAF using federated access control models, i.e. where they authenticate at their home institution. Following successful authentication, depending on their privileges they should be able to access diverse data sets and analyse them according to their research needs as if the data was available directly through the web site (portal) they are accessing. At present over 300 major data sets from a multitude of organisations is made available through the AURIN e-Infrastructure and this number continues to grow. Indeed based on extensive feedback from the research community the primary need of the AURIN e-Infrastructure is to allow access to data.

To deliver this requires that programmatic access to data is achieved, or more specifically federated access to the distributed databases and systems. However at present many data providers, especially national and state-based agencies, do not currently offer programmatic access to their data resources. Rather, many data providers have web sites through which data can be found and accessed via a variety of html/web-based mechanisms, e.g. downloadable Excel spreadsheets or .zip files from the ABS. Being able to access distributed data sets from multiple organisations through a single programmatic interface would greatly simplify the life of many urban researchers and allow major urban and built environment research questions to be tackled.

To understand how the AURIN e-Infrastructure is delivering an Australian urban research gateway, we highlight initial results from some of the early lenses. For each of these we highlight the kinds of data sets that are being made available and illustrate representative use cases demonstrating the utility of the tools that have been provisioned thus far.

IV. AURIN RESEARCH CASE STUDIES

In all of these examples, it is important to emphasise that these are examples of what can be undertaken through the AURIN e-Infrastructure, i.e. the intention here is not to infer specific scientific results based on data that has been used.

A. Population Demographics

There are many research challenges associated with the continued growth and livability of Australian cities. The changing population profiles with an increasingly older generation, the influx of immigrants and their integration into society are some of the challenges facing Australia (and many other countries). These are not just research challenges but broader societal and governmental challenges that must be addressed. AURIN has identified a broad spectrum of data sets [18] and tools that must be incorporated to support research into this area as shown in Figure 4.

As a representative example of the use of the AURIN e-Infrastructure, we consider the city of Sydney and in particular the local government authorities of Sydney. Selecting the situational context through the process illustrated in Figure 3, and searching for data using the interface shown in Figure 2, a reduced (filtered) subset of the AURIN data is accessible.

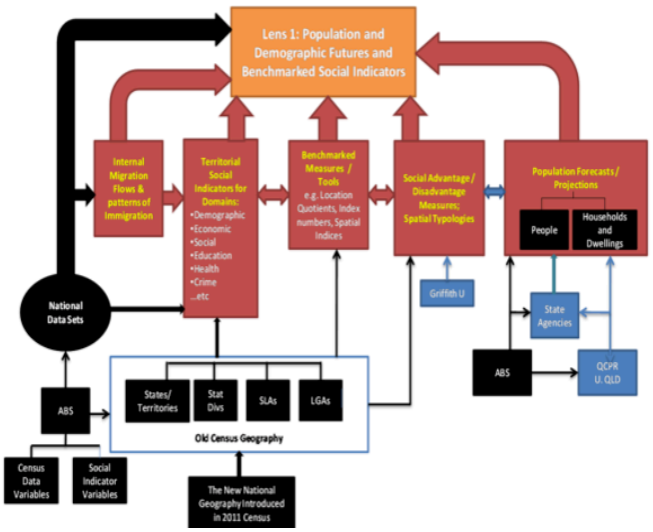

Figure 4: AURIN Demographic Lens Data Landscape

1) Population Demographics for Sydney

In this scenario we focus on the population distribution of individuals living in Sydney according to the 2006 Census; the number of individuals in the labour force, i.e. individuals of a working age, their income levels and their voting patterns. These data sets are accessible from Landgate in Western Australia (https://www2.landgate.wa.gov.au); Centre of Full Employment and Equity (http://e1.newcastle.edu.au/coffee) at the University of Newcastle, New South Wales, and the Australian Election Booth Catchment Areas from the ANDS Spatially Integrated Social Science (SISS) (http://www.itee.uq.edu.au/eresearch/projects/ands/siss) at the University of Queensland.

The population distribution for Sydney is shown in the choropleth map shown in Figure 5 (using a Jenks classifier set to 3 – hence three colour codes). The labour force of Sydney is overlaid on top of the choroplath map as centroids. Finally the LGA voting profiles of Sydney are also illustrated. As shown, the correlation between lower/higher income population in those LGAs and the voting patterns given for the Australian Labour Party from those LGAs.

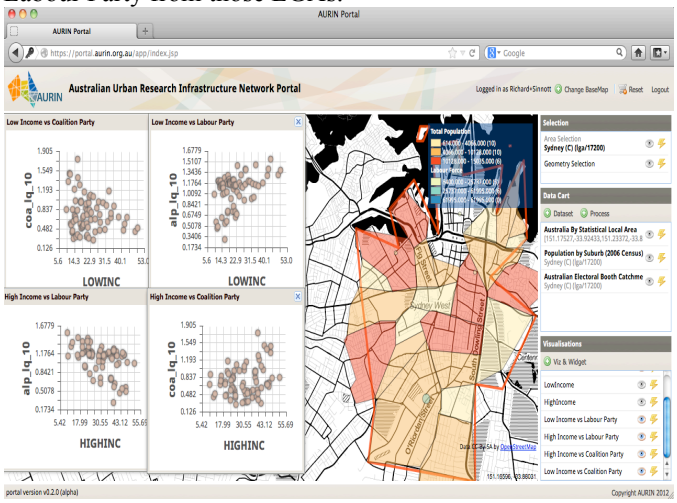

Figure 5: Voting Profiles for Low/High Earners in Sydney

This live access to distributed data and mashing and visualizing is typical of the kinds of functionality that Australian demographic researchers have hitherto not had. Instead they would typically access a wide range of different web sites and download Excel spreadsheets, which would then be imported into statistical tools such as STATA or R. They would also not be able to undertake the advanced geospatial analyses and visual capabilities as shown in Figure 5.

B. Economic Analyses and Urban Labour Markets

Australian cities as with many countries face challenges brought about by increasing population growth and the continued evolution of the global financial crises and the impact on employment and labour in cities. This challenge is further magnified with the increasing trend for longer life spans. Furthermore given the increase in price of houses facing many major cities around Australia, there is a tendency for city growth where workers have to commute increasing distances to/from work. As noted, to tackle such phenomenon, AURIN has identified a broad spectrum of data sets and tools [18] that must be delivered to the wider research community as shown in Figure 6.

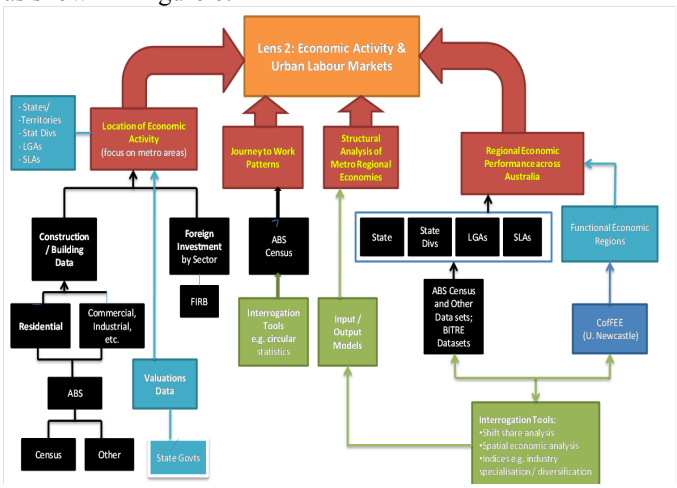

Figure 6: AURIN Socio-economic Lens Data Landscape

1) Employment and Urban Economics for Brisbane

Understanding local and regional employment trends and their impacts on the local economy (and vice versa) is a major factor affecting many cities. How do these local trends compare to the national average is a key barometer to measure. Shift-share is a widely used analytical technique used to identify industries considered to have a comparative advantage in particular areas [14]. The importance of particular industries on the local economy can have a major influence on society, e.g. should that industry suffer economic difficulties.

Brisbane as with many Australian cities has areas with pockets of socio-economic difficulties where local investments and government support are often used to kick start improvements in the local economy. Identifying these

deprived areas and measuring their levels of depravation is a key component of urban economics.

Figure 7 illustrates how such information is accessed and used through the AURIN urban science gateway. Data on socio-economic variables including classification of household income from the University of Queensland compared to the total population are shown in the choropleth map. Also plotted are those statistical local areas with lower weekly income. As indicated by the density of the bar chart, the AURIN platform allows extensive information to be returned and analysed.

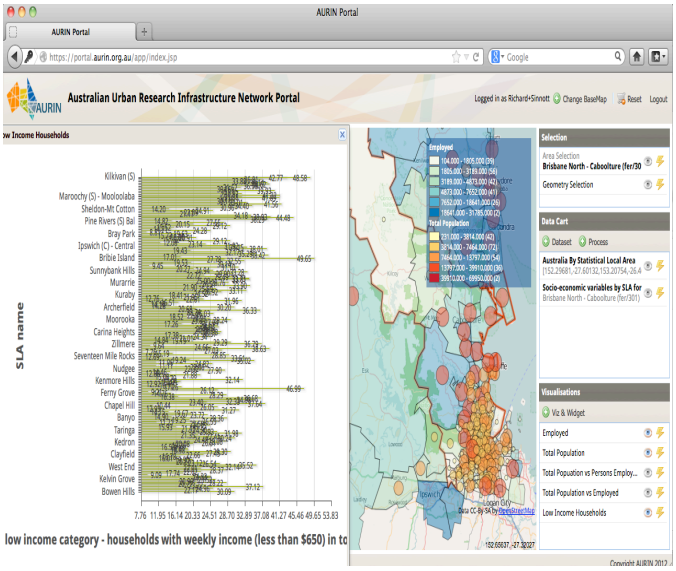

Figure 7: Brisbane Low Income Households and Employment Patterns

C. *Urban Health and Melbourne*

A major challenge facing society is the increased urbanization and its impact on the health and wellbeing of citizens. Living in increasingly populated urban environments has a range of factors that can influence the health of individuals. From the spread of diseases through the increased density and centralisation of the population, the mental health of individuals living in cities, to the increasingly sedentary lifestyle of individuals, where physical activity is decreasingly undertaken. Health data can be specific health information on given individuals with obvious security and privacy considerations that must be addressed. Health data is also often aggregated by agencies for wider research purposes. AURIN deals with both flavours of data from a range of agencies. To tackle such scenarios, the AURIN project has identified [18] a range of tools and data sets that need to be brought into the urban research gateway as shown in Figure 8.

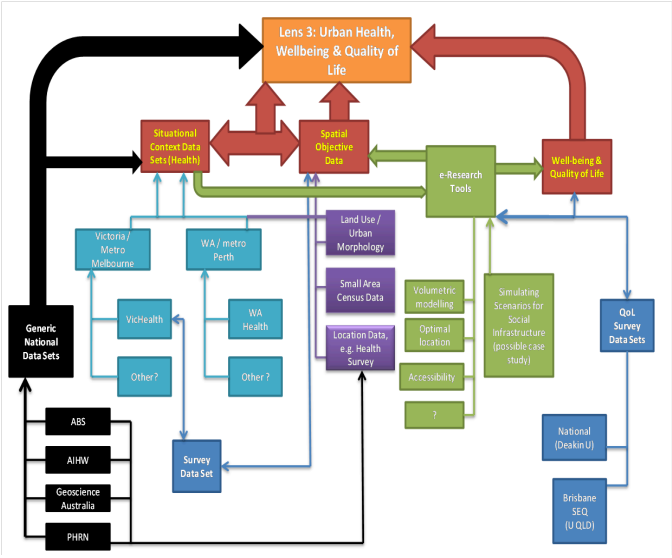

Figure 8: AURIN Health Lens Data Landscape

1) *Health Indicators and Life Expectancy for Melbourne*

To understand how AURIN supports urban health research challenges, we outline a typical research use case linking individual level survey data, e.g. questionnaires, with other data to derive particular health measures. In 2012, the Victorian Department of Health completed a major survey on the health and lifestyle of Victorian residents. This included responses from over 25,000 individuals on a range of questions concerning their health and wellbeing and factors that can influence this, e.g. smoking, alcohol consumption. Access to such individual responses is restricted and subject to strict information governance constraints. These data sets give a representative, statistically relevant snapshot of the Victoria population and cover measures such as "Subjective Wellbeing" and "Work-Life Balance". Complementing these surveys are data from the ABS and PHIDU. The ABS Census gives the most detailed information available for the Australian population covering a variety of aspects of population demographics and living, working in Australia more generally. PHIDU hold a rich collection of data covering births, deaths, health, e.g. cancer screening. At present PHIDU make available over 150 major data sets covering a variety of health related issues across Australia to AURIN.

Figure 9 shows how indicators from VicHealth data can be used to improve understanding of population health survey data. Figure 9 shows the Victoria wide data for those who feel safe walking at night indicator compared with the indicator for those who partake in civic engagement activities (VicHealth 2011 survey). This data covers all of the local government authorities of Victoria and is illustrated through choropleth maps (feeling safe walking at night indicator) and centroids (engagement in civic participation indicator).

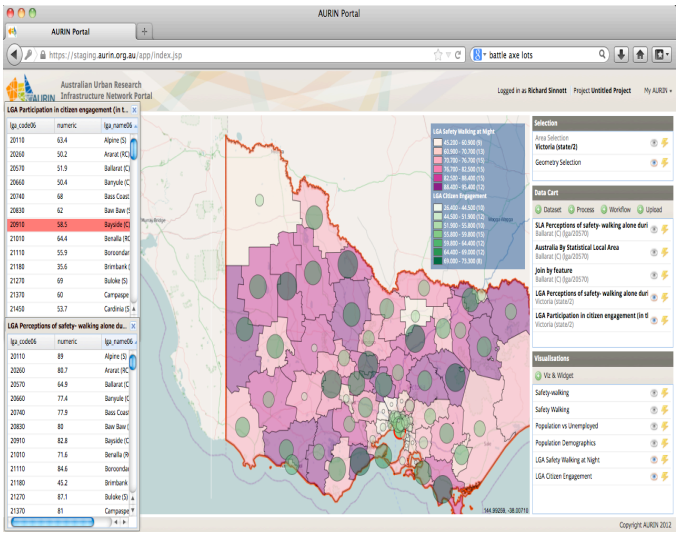

Figure 9: Visualisation of Feeling Safe Walking at Night Indicator and Engagement in Civic Engagement Activities Indicator for all LGAs of Victoria (VicHealth 2011).

This data (these indicators) have been aggregated at the SLA and LGA levels by VicHealth, however work is currently ongoing to utilise the unit level (non-aggregated) data from VicHealth. This is based on the geo-location of the individuals who have participated in the survey. This geo-location allows for a range of analytics to be supported without revealing the identity of the individuals themselves. For example, knowing how many individuals purchased alcohol in the last week may be directly related to how close they live to alcohol selling outlets. Similarly, knowing how little sleep they have might be related to local noise pollution, e.g. living next to major urban transport junctions.

V. Related Work

In many respects the AURIN work is tackling a common research phenomenon. All research disciplines are becoming increasingly driven by the volume of data that can be created and exist in various forms on the Internet [13]. It is the case that almost all research endeavours are limited by the ability to discover, access and optimally use web based data.

To tackle this across Australia, major initiatives have been sponsored. Most notably amongst these are the Australian National Data Service (ANDS – www.ands.org.au) and the $50m Research Data Storage Infrastructure (RDSI – www.rdsi.uq.edu.au) projects. ANDS was largely focused on research data catalogues and especially metadata related to the long term storage and archiving of data. RDSI is to be focused on actual research data itself. Neither of these projects have successfully managed to tackle the heterogeneity of research data integration that typifies what AURIN is doing. This is natural in many respects since they are generic and research domain agonstic.

In the urban and built environment domain there have been a variety of efforts that have looked at aspects of the challenges in supporting data-driven research. The UK ESRC funded Data Management through e-Social Science project (DAMES – www.dames.org.uk) developed a variety of specialised research environments through which a range of distributed social science data sets and associated tools were made available. These covered such as occupational data resources; educational data resources; ethnicity/minority data resources, and e-Health data resources [15]. However the magnitude of the AURIN project and the live access to distributed data is a major enhancement of what was attempted through DAMES.

The National e-Infrastructure for Social Simulation (NeISS – www.neiss.org.uk) project also developed a portfolio of e-Social science solutions that allow researchers to explore a variety of what-if scenarios, using data sets such as the UK Census [10], the British Household Panel Survey combined with real time data such as Twitter. However, this was largely focused on social simulation with a relatively small set of data providers. Again the magnitude of the AURIN undertaking is much more ambitious.

A range of efforts are currently on-going to harmonise international data resources and archives of relevance to urban and built environment researchers. Examples of these include the European Council for European Social Science Data Archives (CESSDA – www.cessda.org) which aims to harmonise social science data archives across Europe, and the EU INSPIRE initiative (www.inspire.jrc.ec.europa.eu) to support global geospatial data initiatives. In the geo-spatial area, the Open Indicators Consortium initiative (www.oicweave.org) aims to develop a visualization platform for any dataset by anyone. This solution currently allows to deploy websites aimed at providing visual exploration capabilities for a specific, locally held dataset in a web based-environment.

The CyberGIS initaitive supported by the NSF (http://cybergis.cigi.uiuc.edu) is perhaps closest to AURIN. While not explicitly aimed at the urban and built environment research disciplines, the aim of exposing computing facilities to process and analyse spatial data may offer collaboration opportunities with AURIN.

It is the case however that the pace of data generation and data availability brought about by the rise in the use of the Internet and associated technologies, e.g. Web 2.0 and social media, has overtaken the way in which researchers themselves are able to discover and utilise the ever expanding volumes of digital data. The AURIN e-Infrastructure has been developed to be generic and to scale with the growth of data, however the data deluge and finding the right data remains a challenge. As one example, there are at present over 300 data sets that are made available through the AURIN e-Infrastructure. Searching for a common urban theme, e.g. "*employment*" will return matches from over 20 organisations. When the e-Infrastructure scales to up to 3000 data sets (each of which can contain up to 200 variables) the magnitude of data management will be seriously challenged. However when compared with searching for "*employment Australia*" which returns over 118 million matches, it is clear that the urban research focus of AURIN is a vast improvement of more generic search engines.

VI. CONCLUSIONS

In this paper we have demonstrated the application of the AURIN urban research gateway in a range of scenarios and illustrated how it directly supports *data-driven* urban research. This work is far from complete and an extensive portfolio of activities for lens-specific projects and their integration into the AURIN e-Infrastructure is very much ongoing. It is expected that the AURIN project will include up to 50 separate lens-specific research subprojects that will be incorporated through 2013 and beyond.

The work and scope of AURIN continues to extend. An increasing focus of AURIN is on incorporation of social media data. Harvesting and use of Twitter data is already supported with tools that allow tracking of the location and movement of tweeters and for example, the languages that they tweet in [2, 16]. Such information provides a different, real time perspective of health information from providers like the ABS, VicHealth and PHIDU.

AURIN is also attempting to provide a degree of intelligence in supporting researchers. This is being achieved in several ways: through repeatable workflows that document the scientific process; through classification and use of variables and their exploitation by tools, e.g. it is not possible to take the average of a categorical variable such as 1/0 for true/false [17]. Importantly, AURIN is allowing researchers to collaborate. This working together and peer review is a key aspect of AURIN. Given the diversity and breadth of the research domains, there is no single expert. Rather multiple experts must collectively work together to tackle the major challenges facing Australian cities and its future as a whole.

Finally we note that the AURIN e-Infrastructure is very much a supporting activity. That is, the work in the e-Infrastructure development is not targeted at delivering novel IT solutions per se nor exploring research challenges in e-Infrastructures, but on supporting the urban research community in *their* research needs. It is worth noting that the implementation work described in this paper commenced in earnest towards the end of 2011 and is now actively being used to convince the varied urban researchers associated with the different lenses, and the associated urban research data stakeholders of the vision of the e-Infrastructure as a whole. The project as a whole is planned to run to mid-2015.

ACKNOWLEDGMENTS

The authors would like to thank the AURIN Technical Committee and Expert Groups that are directly shaping these efforts. The AURIN project is funded through the Australian Education Investment Fund SuperScience initiative. We gratefully acknowledge their support.

REFERENCES

[1] AURIN Final Project Plan, http://aurin.org.au/resources/final-project-plan

[2] R.O. Sinnott, G. Galang, M. Tomko, R. Stimson, *Towards an e-Infrastructure for Urban Research Across Australia*, IEEE e-Science Conference, Stockholm, Sweden, December 2011.

[3] R.O. Sinnott, C. Bayliss, G. Galang, P. Greenwood, G. Koetsier, D. Mannix, L. Morandini, M. Nino-Ruiz, C. Pettit, M. Tomko, M. Sarwar, R. Stimson, W. Voorsluys, I. Widjaja, *A Data-driven Urban Research Environment for Australia*, IEEE e-Science Conference, Chicago USA, October 2012.

[4] M. Tomko, C. Bayliss, G. Galang, P. Greenwood, G. Koetsier, D. Mannix, L. Morandini, M. Nino-Ruiz, C. Pettit, M. Sarwar, R. Stimson, R.O. Sinnott, W. Voorsluys, I. Widjaja, *The Design of a Flexible Web-based Analytical Platform for Urban Research – Systems Paper*, ACM SIGSPATIAL GIS 2012, Redondo Beach, USA, November 2012.

[5] R.O. Sinnott, C. Bayliss, G.Galang, D.Mannix, M. Tomko, *Security Attribute Aggregation Models for e-Research Collaborations*, Proceedings of TrustCom 2012, Liverpool, UK, June 2012.

[6] C. Pettit, R. Stimson, M. Tomko1, R.O. Sinnott, *Building an e-infrastructure to support urban and built environment research in Australia: a lens-centric view*, Surveying & Spatial Sciences Conference 2013, Canberra, Australia, April 2013.

[7] B. Javadi, M. Tomko, R.O. Sinnott, *Decentralized Orchestration of Data-centric Workflows Using the Object Modeling System*, 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2012), Ottawa, Canada, May 2012.

[8] B. Javadi, M. Tomko, R.O. Sinnott, *Decentralised Orchestration of Data-centric Workflows in Cloud Environments*, Future Generation Computing Systems, 2013, http://dx.doi.org/10.1016/j.future.2013.01.008

[9] B. Javadi, R.O. Sinnott, J. Abawajy, *Scheduling of Scientific Workflows in Failure-prone Hybrid Cloud Systems*, ASE Special Issue Journal of IEEE CloudCom-12, February 2013.

[10] M. Birkin, R. Allan, S. Beckhofer, I. Buchan, J. Finch, C. Goble, A. Hudson-Smith, P. Lambert, R. Procter, D. de Roure, R.O. Sinnott, *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).

[11] S. McCafferty, T. Doherty, R.O. Sinnott, J. Watt, *Supporting Research into Depression, Self-Harm and Suicide across Scotland*, Journal of the Philosophical Transactions of the Royal Society A, July 2010, (DOI:10.1098/rsta.2010.0150).

[12] M.S. Sarwar, R.O. Sinnott, T. Doherty, J. Watt, *Towards a Virtual Research Environment for Language and Literature Researchers*, Journal of Future Generation Computer Systems, Elsevier, March 2012, http://dx.doi.org/10.1016/j.future.2012.03.015.

[13] T. Hey, A.Trefethen, *The Data Deluge: An e-Science Perspective*, Grid Computing: Making the Global Infrastructure a Reality (eds F. Berman, G. Fox and T. Hey), (doi: 10.1002/0470867167.ch36)

[14] Shift Share Analysis, http://en.wikipedia.org/wiki/Shift-share_analysis

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

[16] C. Pettit, I. Widjaja, P. Russo, R.O. Sinnott, R. Stimson, M. Tomko, *Visualisations for Exploring Urban Space and Time*, International Society for Photogrammetry and Remote Sensing, Melbourne Australia, September 2012.

[17] S.S. Stevens, *On the Theory and Scales of Measurement,* Science 103 (2684):677-680, 1946.

[18] R. Stimson, M. Tomko, R.O. Sinnott, *The Australian Urban Research Infrastructure Network (AURIN) Initiative: A Platform Offering Data and Tools for Urban and Built Environment Researchers across Australia*, State of Australian Cities, Melbourne, Australia, November 2011.