Towards an e-Infrastructure for Urban Research Across Australia

Richard O. Sinnott[1], Gerson Galang[1], Martin Tomko[2], Robert Stimson[2]

[1]Melbourne eResearch Group,
Department of Computing and Information Systems,
University of Melbourne, Melbourne, Australia
rsinnott@unimelb.edu.au

[2]Department of Architecture Building and Planning
University of Melbourne
Melbourne, Australia

Abstract— **Many challenges facing urban and built environment researchers stem from the complexity and diversity of the urban data landscape. This landscape is typified by multiple independent organizations each holding a variety of heterogeneous data sets of relevance to the urban community. Furthermore, urban research itself is diverse and multi-faceted covering areas as disparate as health, population demographics, logistics, energy and water usage, through to socio-economic indicators associated with communities. The Australian Urban Research Infrastructure Network (AURIN) project (www.aurin.org.au) is tasked with developing an e-Infrastructure through which a range of urban and built environment research areas will be supported. This will be achieved through development and support of a common (underpinning) e-Infrastructure. This paper outlines the requirements and design principles of the e-Infrastructure and how it aims to provide seamless, secure access to diverse, distributed data sets and tools of relevance to the urban research community. We also describe the initial case studies and their implementation that are currently shaping this e-Infrastructure.**

Keywords- Urban Research, e-Social Science, Security

I. INTRODUCTION

Urban environments are both physical and social entities, shaped by natural resources as well as by human behaviour [1]. They are complex adaptive systems, comprising highly interdependent and dynamic elements. These can include the buildings themselves; the open spaces and parkland that exist within them; domestic and commercial property markets; healthcare; water and waste management systems; energy and communications networks; socio-demographic variability and vulnerabilities; transport systems with complex mobility patterns including freight logistics and airports; and labour markets and economic dynamics amongst many other elements. Urban systems also comprise less tangible but crucially important elements such as: patterns of human behaviour; differences in structures of residential provision and ownership; measurable or perceived differentiations in equity and access; differential socio-economic capacities to adaptation and change; and disparate levels of understanding of major drivers of change. The urban research landscape can thus be characterised as being extremely broad and covering a multitude of research interests [2].

One artefact of this heterogeneity is that data is often found in a range of formats and available through a diverse range of agencies and data providers crossing the commercial, government and academic sectors. It is fair to say, that to now, *ad hoc* data access and management solutions have been the primary way in which urban and built environment research has been undertaken. This is typified by collections of Excel spreadsheets and files downloaded from a plethora of data providers. Often these data sets themselves are subject to restrictions on their access and usage, e.g. micro-/unit-level data sets such an individual responses in a census survey. In this context, organisations have established silos of data each with their own heterogeneous access demands and data models that prohibit access and usage by the wider research community. Researchers themselves are subsequently left with enormous swathes of distributed data sets from multiple organisations. Assuming they are aware of the existence of these data sets, they are typically tasked with manual collection and manipulation of these data sets through their own in-house techniques and expertise garnered over extended time periods using a range of statistical analysis tools. A further complexity is that the data itself can take many forms and be at many temporal scales: from archived social science resources curated over extended time periods from organisations such a national statistics agencies, to short lived (ephemeral) data sets captured in real-time, e.g. the current position of cars or people in a given urban environment [3].

In this context, the Australian Urban Research Infrastructure Network (AURIN) project (www.aurin.org.au) has received $20 million to remedy this situation for urban researchers across Australia by supporting the '*establishment of facilities to enhance the understanding of urban resource use and management*'. Specifically, the AURIN project is tasked with providing urban and built environment researchers with 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 of Australia's national settlement system. This includes providing an improved evidence base to inform public policy and inform business operations across urban settlements. This is particularly important given national debates on the implications of population growth, how that will be distributed across urban space, the types of urban environments in which

IEEE computer society

diverse peoples will live, and how the nation can achieve the sustainable development of its cities and towns [4].

e-Research, or e-Science has much to offer researchers in this space: technical solutions for 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 the individual organisations involved; supporting and promoting openness and inter-operability in a coordinated manner whilst tackling 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 researchers currently access and use a variety of data sets and associated tools is also highly desireable. Ultimately, a key barometer of the success of AURIN will be in how the 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.

A. Related Work

The AURIN project is not unique in attempting to tackle the many issues with distributed data sets and tools as the limiting factor in enabling research. Numerous projects have explored the general area of e-Social Science and indeed, the AURIN project fully expects to leverage the insights gained in delivering these projects. The 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 included occupational data resources; educational data resources; ethnicity/minority data resources, and e-Health data resources [5,6]. E-Health examples of the application of the DAMES infrastructure focusing upon mental health, depression and suicide and the broader public health story (including linkage with a range of social science and geospatial data sets) are described in [7].

The National e-Infrastructure for Social Simulation (NeISS – www.neiss.org.uk) project has 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. Examples of the application of the NeISS infrastructure are described in [8,28]. These have built upon earlier systems such as MoSeS [9] and Genesis [10]

Tools such as Transport Analysis and Simulation System (Transims) [11] and the Epidemic Simulation System (EpiSims) [12] allow to simulate urban models at high levels of complexity including tackling human population dynamics and associated social networks in urban environments at a national scale supporting epidemiology and the spread of infectious diseases; computational and behavioral economics and commodity markets [13].

Similarly many 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 CESSDA efforts [14] to harmonise social

science data archives across Europe, and the EU INSPIRE initiative [15] to support global geospatial data initiatives.

In all of these efforts, however 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 to support research, has overtaken the way in which researchers themselves are able to discover and utilise the ever expanding volumes of digital data. Indeed the urban research landscape faces acute challenges in this regard due to its heterogeneity since urban research often requires access to clinical data, social science data, geospatial data, economics data sets etc. Each of these data domains, has its own particular demands and knowledge that is often as isolated as the data sets themselves.

The AURIN project is thus tasked with delivering an e-Infrastructure through which a variety of urban research areas can be supported, and importantly the inter-operability between these research areas to support inter-organisational, inter-disciplinary urban research.

The rest of the paper is structured as follows. Section 2 describes the general urban research requirements on the AURIN e-Infrastructure. Section 3 focuses on the AURIN e-Infrastructure architecture itself and its associated components. Section 4 enumerates some of the key data providers and data sets that this infrastructure needs to make available. Section 5 demonstrates the utility of this e-Infrastructure through exemplar case studies. Finally section 6 draws some conclusions on the work and identifies the plans for the future.

II. AURIN E-INFRASTRUCTURE REQUIREMENTS

The AURIN e-Infrastructure must meet a range of demands from both the urban research communities, and the various data stakeholders involved across Australia. We outline these here since they inform the design of the e-Infrastructure itself.

A. Single sign-on

The AURIN e-Infrastructure needs to support the notion of single sign-on. That is to say that urban and built environment researchers wishing to access distributed, heterogeneous data resources and tools across Australia should be able to do so without the need for multiple authentication and/or authorisation steps, e.g. username and password pairs and/or challenge/responses. This should also be completely transparent to the end users themselves.

B. Autonomy

Data providers should be able to define and enforce their own local discretionary policies on access to and usage of their own local resources. These policies can be expressed in many ways depending upon the procedures and practices of the providers themselves; i.e. technical solutions that they are familiar with. Mandating a particular technology or software stack will not be possible.

C. Usability

The heterogeneous nature of distributed resources and their associated security should be made seamless (transparent) to

the end user researchers themselves. That is, they should be able to access and use data and resources through the AURIN e-Infrastructure with minimal knowledge of the middleware and technical solutions used to deliver this access. For many researchers, the base level of technology needs to be an internet browser.

D. Accountability

In many cases, end users may need to be aware of the obligations that arise with access to and usage of certain data sets from particular providers. Thus, providers of sensitive data need to be completely satisfied that user's who are accessing and using such data as part of their own research, are fully aware of the terms and conditions for access and usage, and consequences for misuse. Capturing access and usage for future auditing is an essential component of the AURIN e-Infrastructure work.

E. Openness and inter-operability

In developing the AURIN e-Infrastructure, it is essential that wherever possible open solutions are applied. Usage of proprietary software, which locks out subsets of the urban research community, e.g. due to licensing issues, should be avoided if possible. Furthermore, open standards and standardised programming interfaces should be exploited and supported leveraging best practice in the domain wherever possible.

F. Collaborative nature

The AURIN e-Infrastructure needs to support multiple research collaborations which may evolve over time. Researchers should be able to contribute in a range of research initiatives though the AURIN e-Infrastructure and input their own thoughts and ideas to its evolution. This includes: the way in which the e-Infrastructure is delivered; the content it makes available; and how the researchers wish to collaborate with one another.

Many of the above general requirements can be addressed through the development and supported of e-Science virtual organisations (VOs) and especially those based around data-driven collaborations [25]. For AURIN, it is important that the VOs are dynamic and evolving, and not simply offer/support static environments giving access to hard-coded data sets and resources. Researchers will use the VO resources as the source of information, e.g. to access data, as well as feed the VO with data sets themselves, e.g. results and/or comments on data they have analysed and visualized through the e-Infrastructure.

Another way in which the AURIN e-Infrastructure can thus be considered is that it should support a collection of VOs on a common infrastructure, where services and targeted data sets of relevance to given urban research communities in those specific areas are made available.

III. AURIN E-INFRASTRUCTURE DESIGN

A simplified view of the overall architecture for AURIN is depicted in Figure 1. The model is based around a unified research environment for a spectrum of urban and built environment research scenarios and researchers. For the purpose of illustration, the focus in the figure and in the ensuing discussion is on energy and water (EW) research as an example, but the discussions are generic and can be used for urban transport or other research scenarios, and indeed the interworking of these services/environments.
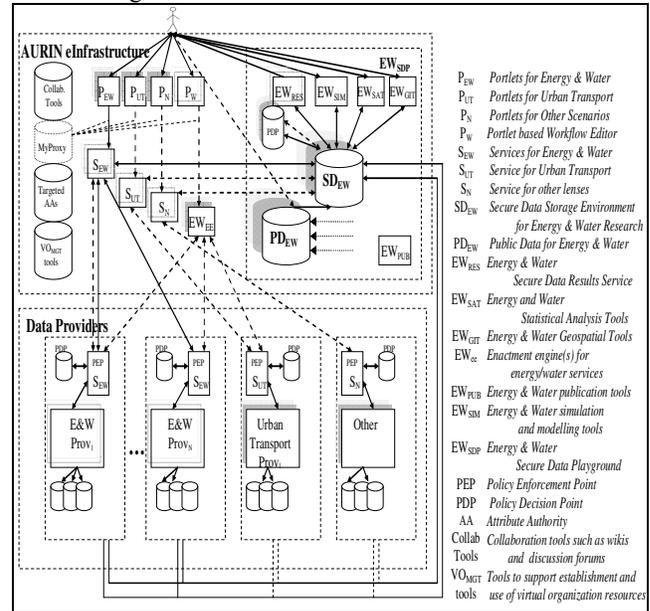

Figure 1: Simplified model of the AURIN e-Infrastructure

The AURIN e-Infrastructure is accessible through a LifeRay-based portal (www.liferay.com). LifeRay provides a platform and framework for developing collaborative and social applications and services. It provides wiki and discussion forum support out-of-the-box. LifeRay also supports the latest portlet specification: JSR-286. It offers built-in support to a number of authentication and authorization system including CAS (Central Authentication Service), NTLM (NT Lan Manager), OpenID, and OpenSSO. By using LifeRay, we have been able to leverage a number of the portal services it provides directly. The AURIN portal itself has been deployed within (and made available) through the Australian Access Federation (www.aaf.edu.au) to support federated authentication.

The e-Infrastructure architecture has adopted a typical n-tier service-oriented architecture, albeit one with specific focus on ensuring both security of the data and supporting research collaboration. This service-oriented architecture encompasses both traditional Simple Object Access Protocol (SOAP)-based web services and Representational State Transfer (REST) based web services accessible over *http*.

The data sets themselves are accessed, used and manipulated through targeted JSR-286 compliant portlets which interact with a range of distributed services (S_{EW}) developed and deployed as part of the e-Infrastructure. Many of the AURIN services will leverage hosting services offered through distributed partner sites across Australia (EW_{Prov}) and these will (potentially!) have their own access and usage

policies in place, realized through appropriate 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 [16-19] 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. This attribute is used to persistently identify users throughout the system, i.e. in the portal and when accessing remote resources. The authorization decision can itself be made through pulling of additional authorization (X509 attribute certificates) from an AURIN-specific (trusted) attribute repository (LDAP server), or through direct local access control decisions based on knowing the identity of the individual making the request. 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 [20,29]. 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, i.e. an IdP specifically set up for AURIN that can be used to authenticate and authorise individuals from organizations not involved in the AAF directly.

Digitally signed user privileges (roles) either delivered directly through the AAF as X509 attribute certificates or maintained locally and coupled with the federated user identity are used to provide users with access to targeted portlets for their own research areas of interest, i.e. to create a security-oriented, tailored research environment offering tools and data sets reflecting their needs, interests and importantly, their assigned AURIN privileges.

The actual data access and delivery in AURIN is based upon agreed interfaces to known data sets and services. Typically, this is realized through a pre-agreed interface (portlet) where collections of variables can be selected. The queries that are submitted through these portlets may be sent to locally supported services as part of the n-tier architecture where business / research-specific logic is applied before they are sent onwards to the energy and water service providers, or they can be sent directly to remote services giving access to data. A variety of such services for energy and water providers (S_{EW}), urban transport data providers (S_{UT}) or indeed other service/data providers (S_N) will exist depending on the nature of the research question/implementation streams for AURIN.

In short, when different users log-in to the AURIN portal environment through the AAF they are presented with different *potential* views of data sets that may be made available by data providers. It is emphasized that this is a view of variables that can be selected and not a view of data itself. That is, these portlets offer access to a collection of variables whose values, i.e. the data itself, may or may not be released to the requestors depending upon the data access and usage policies of the data providers involved.

For security-oriented scenarios, depending on the nature and source of the data, it may be the case that data provider data

sets are not directly (immediately) accessible to the end users through the portal. In the most conservative case, they will only be access/analysed through a predefined collection of targeted 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; e.g., recoding routines to take local geospatial data sets (postcodes, boundary data sets etc) from results data sets and map these onto larger local authority level coordinates, or for recoding specific variables according to particular health classifications.

Data playgrounds have been identified as 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}), e.g. by mirroring of associated data from known energy and water data providers or uploaded by members of the research community. Data playgrounds themselves can be realised 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 (subject to security constraints for example) support larger-scale data processing capabilities; e.g., 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 architecture proposed, 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, e.g. they can be linked and joined, this linkage information can subsequently be removed to avoid data disclosure risks [26,27]. 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 individualised data sets. Each of these data playgrounds will have their own security constraints and access and usage policies.

It has been recognised 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 co-ordinate 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 [21]. 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 [22].

IV. AURIN RESEARCH DATA LANDSCAPE

There are a huge array of organizations and data sets that currently exist across Australia that are fundamental to supporting urban research. 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 recognized 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 (*lenses*) 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. The initial set of AURIN lenses that have been identified include:

- Population and demographic futures and benchmarked social indicators;
- Economic activity and urban labour markets;
- Urban health, well-being and quality of life;
- Urban housing;
- Urban transport;
- Energy and water supply and consumption;
- City logistics;
- Urban vulnerability and risks;
- Urban governance, policy and management;
- Innovative urban design.

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 areas are themselves inter-related. 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.

There are numerous data providers that hold key data sets relevant to many of these lenses. At a national level, these include major governmental organizations such as: the ABS; Geoscience Australia (www.ga.gov.au); the Bureau of Infrastructure, Transport and Regional Economics (www.btre.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.

There are also numerous State-based organizations (Victoria, New South Wales etc) that hold key data sets. These include local government agencies and authorities including transport agencies (VicRoads - www.vicroads.vic.gov.au/), health agencies (VicHealth - www.vichealth.vic.gov.au/) amongst many others. Numerous commercial organizations also hold data sets that need to be unlocked for urban researchers, e.g. energy and water suppliers, real estate companies etc. Similarly, a wide array of research projects also exists covering aspects of the urban research landscape, e.g. the $100m Cooperative Research Centre for Spatial Information (CRC-SI) project (www.crcsi.com.au).

It is the case that most of these data providers do not currently offer programmatic access to their data resources, i.e. through web services that can be invoked. Rather, many of the data providers have web sites through which data can be found and accessed via a variety of html/web-based mechanisms. Typically these are Excel spreadsheets or .zip files containing a range of data. As one example, the 2006 labour force survey of the ABS contains 23 separate tables covering labour force tables for each State; for labour force status by sex; aggregate monthly hours worked etc. Being able to access these data sets through a single programmatic interface would thus simplify the life of many urban researchers.

The AURIN project is working with all of the aforementioned organizations in helping to support programmatic access to subsets of their data of relevance to the different lenses. In the case of the ABS for example, REST-based services are in the process of being deployed which will allow, for the first time, direct (programmatic) invocations of services providing live access to ABS data.

The vast majority of these data sets are geospatial coded in some form. The project aims to leverage best practice in access to and usage/linkage of geospatial data. The Geospatial community at large through efforts of organizations such as the Open Geospatial Consortium (OGC – www.opengeospatial.org) has defined a collection of core services through which geospatial data can be accessed, used and manipulated. The services of relevance to AURIN include:

- Web Feature Service (WFS) which provides interfaces for data manipulation operations of geographic features, e.g. to get or query features based on spatial and non-spatial constraints;

- Web Map Service (WMS) which allow for creation of maps (images) of spatial data;
- Web Coverage Services (WCS) which allow requests for geographical coverages, e.g. space-time varying phenomenon in a given geographical area;
- Web Processing Services (WPS) which provide an interface that facilitates the publishing of geospatial processes, e.g. services that support calculations, and allows clients to discover and use those processes. A WPS supports simultaneous exposure of processes via web service (SOAP) and REST-based services;
- Web Perspective Services allowing for the visualization of 3D data and landscape settings.

To support the geospatial needs of the AURIN project, the e-Infrastructure has established its own GeoServer instance (www.geoserver.org) which provides WFS, WMS and WCS capabilities and is used to integrate and display a variety of geospatially coded data sets of relevance to urban researchers.

V.   CASE STUDIES

The initial stages of the AURIN implementation have focused on the delivery of the core capabilities for researchers to discover and select data; for spatial and attribute-based filtering; for data-driven linkage and basic selection; for data classification and sorting, and for cartographic visualization and charting of analysed data sets. Such capabilities underpin urban research across all lenses.

The initial AURIN lenses are focused on datasets and functionalities requiring the analysis of phenomena at a relatively small-scale, across national datasets of a highly aggregate nature. Underpinning such analyses is a hierarchy of administrative boundaries to which statistical datasets are related. The ABS is a primary source of such data in Australia. Geospatial datasets together with the boundary geometries are accessible through a spatial relationally database (PostGIS) available at the back-end of the portal. This database includes a spatial index assuring fast data access to national urban researchers. This local storage supports highly static geometries which allows for a speed-up of the responsiveness of the system, avoiding costly processing of verbose Geography Markup Language-encoded (GML) geometries [23].

The geospatial interface to the AURIN portal allows users to drill into national, regional and local/suburban urban data sets. This interface exploits a geographic Gazetteer (a place-name lookup service), allowing directed adaptive zooming to a location of choice as displayed in Figure 2.
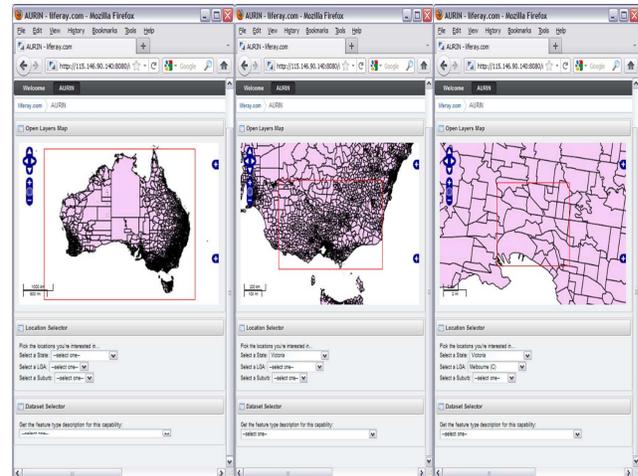
Figure 2: Drilling into Australian Geospatial Data

The portal currently makes available a rich variety of urban data sets from across Australia. These include the Landgate resource from Western Australia (www.landgate.wa.gov.au) which makes a variety of data available through OGC-compliant WFSs. The actual WFS invocation itself is targeted to the specific region selected in the portal, thereby reducing the total data sets that are returned.

The datasets that are currently available through the portal include amongst others: a wide array of employment statistics; population statistics; household income; and language ability information across Australia. It is important to note that these data sets are available through federated (remote) services, and are not stored locally within the portal for example.

To understand how these data sets can be used we consider two typical scenarios of relevance to urban researchers:

- *the correlation between the number of people who don't speak English fluently and their associated unemployment rate across Melbourne;*
- *the correlation between the income and the population statistics of the suburbs of Sydney.*

To address the first scenario we select those areas where there is a high concentration of non-native English speakers. This data is available through the 2006 Census Language by Suburb statistics. The unemployment information per suburb is available through the Employment Statistics from the 2006 Census. Both of these data sets can be linked through geospatial coordinate systems associated with the local authority boundary.

Figure 3: Unemployment vs English Ability for Melbourne

A variety of statistical tools can be used for analysis of such data in an on-line virtual environment, e.g. R, STATA and functions/routines that allow for standard data manipulations. At present the portal simply displays the results of such queries directly to the end users in the portal, however work is now on-going to support the deposition of these data sets into the secure data playground for collaborative analysis and commentary. Similarly integration of analysis tools is an AURIN activity currently being explored.

For the second scenario, having selected the location of interest (Sydney), the user is able to select variables related to median weekly income from the 2006 Census Households by Suburbs survey, and demographic (population) distribution data from the 2006 Population Statistics. The results of accessing and comparing these variables together with charts outlining the suburbs, their populations and the median weekly income is shown in Figure 4.

Figure 4: Suburb Population vs Income around Sydney

In both of these scenarios, selection and linkage of data sets is based on unique identifiers of the records (which can be specified by the user), or based on a spatial join, e.g. the specific suburb code. As shown, the data sets that are returned are displayed in a tabular view, which can be exported, or sorted and filtered based on simple SELECT-like queries. The results of the selection can then be displayed in a cartogram in the same map view as used for the spatial filter. In this example, a classification using the Jenks natural breaks algorithm minimizing in-class variance is shown [24].

The returned data sets can be visualized in a variety of ways, e.g. a bar chart as shown here with user specified x- and y-axis based on the variables of interest.

This initial interaction model shows the basic functionality (workbench) through which users can access and use a variety of distributed data sets in a single unified framework. Work is currently on-going to extend these scenarios with a workflow enactment environment, allowing the chaining of multiple services through the portal giving access to the data sets themselves, and the way they can be manipulated/analysed through targeted tools and services.

VI. CONCLUSIONS

In this paper we have described the on-going work in the AURIN project. This project is very much in the initial phases of development, but proof of concept implementations that are driving forward the efforts of the e-Infrastructure and the associated urban research lenses has already been demonstrated. Indeed the project as a whole is based upon an agile software development plan. Rapid prototyping to establish the proof of concept and its evolution to meet the specific needs of the research community is at the heart of the implementation work.

The AURIN project is also running in the context of many major e-Infrastructure investment activities that are currently taking place across Australia (in addition to the AAF). These include the $50m Research Data Storage Infrastructure (RDSI – www.rdsi.uq.edu.au) which has a specific focus on supporting storage of nationally significant research data sets, and the $47m National eResearch Collaboration Tools and Resources (NeCTAR – www.nectar.org.au) project which has a specific focus on eResearch tools, collaborative research environments and Cloud infrastructures. The AURIN project is currently engaging directly with these projects in delivery of much of its underpinning infrastructure. For example, the portal and associated services and data services are available on virtual machines made available through NeCTAR. This arrangement will allow the project to scale accordingly when the urban research base demand rises. We fully expect to work with RDSI in development and delivery of the secure data playgrounds, accessible through the AAF. IN short, the AURIN project is being used to help shape major national e-Infrastructure initiatives that are taking place across Australia.

We note also that the work with the AAF is also novel and helping to shape many activities across Australia where finer-grained access control (authorisation) is required. At present the AAF has predominantly been used for authentication

purposes only. Expertise in security policy definition and enforcement is needed for many domains – not just the urban research domain.

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 or 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 May 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 runs to mid-2014.

REFERENCES

[1] M. Freire, R.E. Stren, *The Challenge of Urban Government: Policies and Practices*, World Bank Institute, ISBN 0-8213-4738-1, January 2001.

[2] S. E. Gill and J. F. Handley and A.R. Ennos and S. Pauleit and N. Theuray and S. J. Lindley, *Characterising the urban environment of UK cities and towns: A template for landscape planning*, Journal of Landscape and Urban Planning, Vol 87-3, pp 210-222, 2008.

[3] W. Jepson , R. Liggett , S. Friedman, *An Environment for Real-Time Urban Simulation*, Proceedings of the ACM Symposium on Interactive 3D Graphics, New York, 1995.

[4] I. Correa-Velez, S. M Gifford and S. J. Bice1, *Australian health policy on access to medical care for refugees and asylum seekers*, Journal of Australia and New Zealand Health Policy 2:23, 2005.

[5] P. Lambert, L. Tan, K. Turner, V. Gayle, R.O. Sinnott, K. Prandy, *Data Curation Standards and Social Science Occupational Information Resources*, International Journal of Digital Curation, November 2007.

[6] 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.

[7] 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).

[8] 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).

[9] M.H. Birkin, A. Turner, B. Wu, A. Townend, J. Arshad, J. Xu., *MoSeS: A Grid-enabled Spatial Decision Support System*. Social Science Computer Review, 27, 493-508, 2009.

[10] Genesis Project, http://www.genesis-fp7.eu/

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

[12] C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, M.V. Marathe, *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.

[13] C. L. Barrett, S. Eubank, V.S. A. Kumar, and M. V. Marathe*, The Mathematics of Networks Understanding Large-Scale Social and Infrastructure Networks: A Simulation-Based Approach*, SIAM News, Volume 37, Number 4, May 2004.

[14] Council of European Social Science Data Archives (CESSDA), www.cessda.org

[15] EU Commission INSPIRE, http://inspire.jrc.ec.europa.eu

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

[17] R.O. Sinnott, D. Chadwick, T. Doherty, D. Martin, A. Stell, G. Stewart, L. Su, J. Watt, *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 2008, Lyon, France.

[18] J. Watt, R.O. Sinnott, *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 2011.

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

[20] R. Alfieri, R. Cecchini, V. Ciaschini, L. Dell'Agnello, VOMS, an authorization system for virtual organizations, LNCS 2970, p33-40, 2004.

[21] P. Kacsuka, T. Kiss, G. Sipos, *Solving the Grid Interoperability Problem by P-GRADE Portal at Workflow level*, Future Generation Computer Systems, Volume 24, Issue 7, July 2008.

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

[23] S. Cox, P. Daisy, R. Lake, C. Portele, A. Whiteside, *OpenGIS Geography Markup Language* (GML3.1), Implementation Specification Version 3.1.0, Recommendation Paper, OGC Doc. No. 03-105r1, 2004.

[24] G.F. Jenks, *The Data Model Concept in Statistical Mapping*, International Yearbook of Cartography 7: 186-190, 1967.

[25] I. Foster , C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*, International Journal of High Performance Computing Applications, vol. 0103, 2001.

[26] R.O. Sinnott, O. Ajayi, A.J. Stell, A. Young, *Towards a Virtual Anonymisation Grid for Unified Access to Remote Clinical Data*, Proceedings of the 6th International HealthGrid conference, Chicago, USA, June 2008.

[27] A.J. Stell, R.O. Sinnott, O. Ajayi, J. Jiang, *Designing Privacy for a Scalable Electronic Healthcare Linkage System*, IEEE International Conference on Information Privacy, Security, Risk and Trust (PASSAT 2009), Vancouver, Canada, August 2009.

[28] N. Malleson, M. Birkin, *Towards Victim-Oriented Crime Modelling in a Social Science e-Infrastructure*. Philosophical Transactions of the Royal Society A, July 2011.

[29] Internet2 Grouper Groups Management Toolkit, www.internet2.edu/grouper/