19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Towards an e-infrastructure for urban research across Australia

Martin Tomko [a], **Gerson C. Galang** [a], **Robert J. Stimson** [a] **and Richard O. Sinnott** [a]

[a] *Australian Urban Research Infrastructure Network, The University of Melbourne, 3052 VIC*
Email: tomkom@unimelb.edu.au

Abstract: Australian urban and built environment research covers a multitude of research disciplines investigating social, economic and physical phenomena at a multitude of spatial and temporal scales and across diverse aggregation levels, from individual-level through to cohorts and populations, and across a range of scenarios, e.g. public health, voting patterns. The development of a common software platform (e-Infrastructure) to meet the needs of such research communities involves tackling many challenges associated with data intensive areas of research. This includes dealing with data sets from a multitude of federal, state, municipal, academic and private institutions, which hold vast arrays of heterogeneous data. For many researchers these data sets are difficult to discover, access, interrogate and use more generally. It is also unrealistic to expect researchers to always have the technical capability and capacity to handle such large amounts of data, or to develop data processing tools making use of such data sets, or be able to run computationally intensive simulations and models based on these data sets. Islands of expertise and islands (silos) of data currently exist that have fragmented urban research and thwarted a holistic approach to the study of the Australian urban and built environment system.

The Australian Urban Research Infrastructure Network (AURIN - www.aurin.org.au) is a $20m SuperScience initiative established across Australia that seeks to address this directly by creating an e-Infrastructure aiming at bridging these gaps, by allowing researchers to conduct collaborative research in a security-enabled, browser-based environment providing seamless and transparent access to the distributed data and computational resources across Australia. These include metadata services, federated datasets, data integration and interrogation services, together with advanced visualization, collaboration and data storage capabilities. The goal is to provide access to rich datasets, state-of-the-art data processing tools, as well as a knowledge base where good research practices can be followed and used to assist researchers when navigating through vast data holdings to couple appropriate data and analytical tools for a range of urban research endeavors.

In this paper, we address the fundamental question behind the establishment of this infrastructure, i.e. how to design a versatile and flexible software platform (e-Infrastructure) for urban research? The approach described is centered on establishment of a range of demonstrator projects addressing specific urban and built environment themes and the challenges they give rise to through a common e-Infrastructure. We believe that such an approach will allow delivery of early functionality in supporting a range of urban research scenarios, and at the same time support novel links between tools traditionally not applied beyond individual research fields. Through a common (core) e-Infrastructure, we expect to develop urban research capability that will offer a step change in how urban research is currently conducted, to support multi- and inter-disciplinary research scenarios whilst preserving full functionality in individual urban research strands.

We describe the initial design stages of the AURIN e-Infrastructure from a technical perspective. The approach chosen is based on past experiences from a variety of eResearch initiatives, such as the UK e-Science National e-Infrastructure for Social Simulation (NeISS – www.neiss.org.au) project (Birkin et al., 2010); the Data Management through e-Social Science (DAMES – www.dames.org.au) project (McCafferty et al., 2009); the Spatial Information Software Stack (SISS) eResearch Facility (Liao et *al*., 2009), and the workflow-based image annotation using geographic information retrieval of TRIPOD (Purves et *al*., 2010). We illustrate the utility of the approach taken based on an initial set of demonstration projects exploring demographic and economic investigations of the Australian urban system.

Keywords: *Urban research, eResearch, e-Infrastructure*

1. INTRODUCTION

Urban environments are both physical and social entities, shaped by natural resources as well as by human behaviour, resulting in complex adaptive systems (Freire and Stren, 2001). The elements of urban environments include buildings with different purposes, such as domestic and commercial, the open spaces and parkland that exist between them; transport systems with complex mobility patterns including freight logistics and airports, and others. Urban systems also comprise less tangible but crucially important elements such as patterns of human behaviour; differences in structures of residential provision and ownership; and measurable or perceived differentiations in equity and access.

One artefact of this heterogeneity is that data are often available from diverse agencies and data providers, in a range of formats, and subject to restrictions on their access and usage, e.g. micro-/unit-level data sets such as individual responses in a census survey. So far, ad hoc data access agreements and management solutions have dominated the built environment research domain. This came at a high personal, time and organisational cost to the data providers and researchers alike. Assuming the researchers are aware of the existence of the data sets, they are typically tasked with their manual manipulation and fusion through custom techniques. The undocumented and data handling workflows often result in unexpected or spurious results and a generally low level of replicability of urban research data analysis.

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 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 diverse peoples will live, and how the nation can achieve the sustainable development of its cities and towns (Correa-Velez et *al.*, 2005).

In this paper, we address the fundamental questions behind the establishment of this infrastructure. How to design a versatile and flexible software platform (e-Infrastructure) for urban research? The approach described is centered on establishment of a range of demonstrator projects addressing specific urban and built environment themes and the challenges they give rise to through a common e-Infrastructure. We describe the initial design stages of the AURIN e-Infrastructure from a technical perspective. We illustrate the utility of the approach taken based on an initial set of demonstration projects exploring demographic and economic investigations of the Australian urban system.

The rest of the paper is structured as follows. Section 2 provides a brief overview of the urban research landscape in the context of e-Research. Section 3 describes the general urban research requirements on the AURIN e-Infrastructure and its design principles, capabilities and the key databases envisaged. Section 4 focuses on the currently developed AURIN e-Infrastructure architecture itself and its associated components. Section 5 demonstrates the utility of this e-Infrastructure in exemplar case studies. Finally section 6 draws some conclusions on the work and identifies the plans for the future.

2. 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. e-Research, or e-Science has much to offer to solve the problems noted, faced by urban researchers: 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 a variety of data sets are accessed and interpreted is also highly desirable.

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 (Lambert et al., 2007; Tan et al., 2009). 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 (McCafferty et al., 2010). The Project TRIPOD developed a workflow environment allowing for the synthesis of knowledge and its verbalisation in keywords and image captions, based on heterogeneous spatial datasets and a structured, semantic thesaurus in a federated computing environment (Purves et *al*., 2010).

Tools such as Transport Analysis and Simulation System (Transims) (Nagel et *al.*, 1999) and the Epidemic Simulation System (EpiSims) (Barrett et *al.*, 2008) allow users 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 (Barrett et *al.*, 2004). 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, especially in the domain of crime analysis and simulation, using data sets such as the UK Census (Birkin et *al.*, 2010; Malleson and Birkin, 2011). These have built upon earlier systems such as MoSeS (Birkin et *al.*,2009) and GENESIS (www.genesis-fp7.eu).

The Spatially Integrated Social Science (SISS) project developed an eResearch Facility (Liao et *al*., 2009) allowing for the advanced analysis of various spatially embedded social phenomena in Australia, such as the distribution and variation of voting patterns. This facility has seamlessly linked aggregated spatial datasets and integrated them with advanced spatial analysis and regression tools, thus enforcing good research practice and more intuitive discovery of intricate social patterns.

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 social media has overtaken the researchers' capability to discover and utilise the ever expanding volumes of digital data computationally. The AURIN project is 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.

3. AURIN E-INFRASTRUCTURE REQUIREMENTS AND DESIGN PRINCIPLES

3.1. Core 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. These are translated into a set of critical design principles underpinning the e-Infrastructure's development:

- **Single sign-on:** The AURIN e-Infrastructure needs to support the notion of single sign-on, by supporting the user's own institution's access credentials. 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, using specialised username and password pairs. Instead, their own institution's credentials will be used once, at the start of the session.
- **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, based on their own technical solution – AURIN can not, and will not mandate the technology data providers use to secure their infrastructure and data holdings.
- **Usability:** The heterogeneous nature of distributed resources and their associated security should be made seamless to the researchers themselves. 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, requiring only the ability to operate a Web broser.
- **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.
- **Openness and inter-operability:** It is essential that wherever possible open solutions are applied, avoiding the need for proprietary software installations or licences. Open standards, standardised programming interfaces and commons-like data licencing will be encouraged and supported.
- **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.

3.2. AURIN e-Infrastructure Key Capabilities

The above requirements have been translated into a set of key technical aspects on which the AURIN infrastructure is being built. By guaranteeing support for these key capabilities, AURIN assures that the evolution of the project and the increase of the infrastructure's complexity will not result in a rigid, inflexible and monolithical system.

- **Virtual Organisation Support:** Many of the above general requirements can be addressed through the development and support of e-Science virtual organisations (VOs) and especially those based around data-driven collaborations (Foster et al., 2001). For AURIN, it is important that the VOs are dynamic and evolving, and not simply support static environments giving access to hard-coded data sets and resources. Researchers will use the VO resources as the source of information, and in turn contribute data back to the VOs, based on the analysis undertaken.
- **Secure and Standard Portal-based Collaborative Framework:** The core platform supporting the AURIN e-Infrastructure is a portal framework supporting the latest portlet specification: JSR-286. The AURIN portal itself will be connected to, and will be made available through the Australian Access Federation (www.aaf.edu.au) to support federated authentication. The underlying service-oriented architecture provides functionality to the portlet-based graphical front-end of the AURIN portal, based on federated resources using Simple Object Access Protocol (SOAP)-based web services and Representational State Transfer (REST) based web services accessible over http.**Secure Access to Federated Data and Data Playgrounds:** The actual data access and delivery in AURIN is based upon agreed interfaces to known data sets and services. 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 data service providers, or they can be sent directly to remote services giving access to data. For security-oriented scenarios, data provider data sets are not directly accessible to the end users through the portal. They may only be analysed through a predefined collection of targeted tools offering well specified interfaces through which data processing and analysis can take place. Users may only be presented with variables available in a given dataset, and never with the underlying data themselves. Only the results of a specified analysis are then returned to the user, for instance in the form of a map or graphic.

  Data playgrounds have been identified as a key enabler for this kind of scenario to accommodate data providers wary of 'giving their data away'. Data playgrounds themselves can be realised as a secure file system accessible only to researchers and data providers with sufficient access and usage privileges, or as a DBMS reserved for researchers in a given domain.
- **Data Interrogation Services Orchestrated in Workflows:** Supporting the coupling of the interactions between collections of services and data movement between services and the secure data playground has been identified as an important requirement from the research community. Support for targeted enactment engines that allow, for example, AURIN workflows to be defined, enacted (in isolation and in batch, potentially with varying input variables), logged and shared by other researchers, is a key goal of the AURIN e-Infrastructure. Several workflow environments currently exist, e.g., TAVERNA (http://www.taverna.org.uk/), KEPLER (https://kepler-project.org/), and geospatial standards allowing for web processing services chaining are standardised in the geospatial area (Schut, 2007). Work is currently on-going in evaluating these environments, their logging capabilities, and especially 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).

3.3. Key Data Providers and Data Access

One of the most valued contributions of AURIN is the streamlined access to datasets with, as much as possible, national coverage. Such datasets are only available from governmental agencies and private providers, and the complicated access to these datasets has been a major reason for reduced research flexibility and collaboration in the urban research field.

Many of these data providers have data directly accessible on the Web with variously formatted files 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. At the heart of the AURIN work is providing programmatic access to these data sets. The project has identified a key set of strategic research areas to be realized through implementation streams (lenses) of importance to the urban research community (for a full list of the 10 aspirational lenses, see Section 7.5.3 of AURIN, 2011). Each of these specified their own data requirements, and acted as a VO. Data providers that hold key data sets relevant to the lenses include the ABS; Geoscience Australia (www.ga.gov.au); the Bureau of Infrastructure, Transport and Regional Economics (www.btre.gov.au); numerous State-based organizations, e.g. including transport agencies (VicRoads - www.vicroads.vic.gov.au/); health agencies (VicHealth - www.vichealth.vic.gov.au/) ; private companies such as the Public Sector Mapping Agency (www.psma.com.au); and academic organizations. The AURIN project is working with such organizations in helping to support programmatic access to subsets of their data of relevance to the different lenses. The project aims to leverage best practice in access to and usage/linkage of geospatial and statistical data, such as the Open Geospatial Consortium (OGC – www.opengeospatial.org) and the Statistical Data and Metadata Exchange (SDMX, http://sdmx.org/) standard developed by the IMF, Eurostat, OECD, the ABS and others for access to statistical data sources.

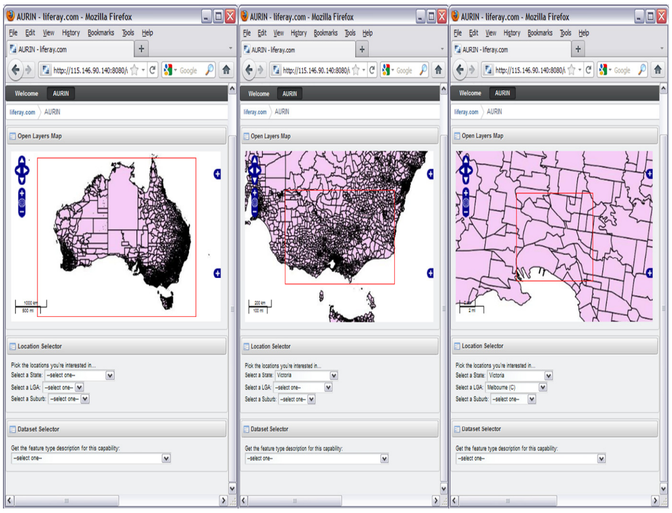
4. THE EMERGING AURIN E-INFRASTRUCTURE IMPLEMENTATION

The architecture of AURIN is completely operationalised based on open source technologies. It is hosted on a multitude of 64-bit Ubuntu Virtual Machines, with a deployment over a GRID-based environment envisaged. Thus far, the portal is implemented using LifeRay (www.liferay.com), and all components are in parallel tested on SAKAI (http://sakaiproject.org/). LifeRay provides an Open Source platform for developing collaborative and social applications and services, satisfying the core requirements, providing out-of-the-box support for wiki and discussion forums, offering built-in support to a number of authentication and authorization systems including CAS (Central Authentication Service), NTLM (NT LAN Manager), OpenID, and OpenSSO. This allows for the use of different authentication systems required by the individual data providers.

Many data providers extensively use services compliant with the OGC data access standards, in particular the Web Map Service and Web Feature Service specifications. In order to be able to seamlessly integrate these services with the portal-based frontend, a cascaded architecture based on Geoserver 2.1.1 has been set up. Geoserver offers a few distinct advantages over other offerings, The mapping front-end is based on OpenLayers, and the graphing engine currently used is JFreeChart 1.0.13. Additional middleware functionality is based on GeoTools 2.7.1, the Java Spring framework and backed by a MAVEN dependency management system. The DBMS currently backing AURIN is Postgres with the spatial plugin PostGIS 1.5.2. The PostGIS database also takes care of a dataset registration system, where the level of geography to which a dataset relates can be stored. The possibility of connecting with a metadata service operating from the Office of Spatial Data Management (OSDM), based on GeoNetwork is currently being developed.

5. THE AURIN CASE STUDIES

The initial stages of the AURIN e-Infrastructure implementation have focused on the core capabilities for data discovery and selection; for spatial and attribute-based filtering; for data-driven linkage, classification and sorting, and for cartographic visualization and charting of the manipulated data subsets. Such capabilities underpin urban research across all lenses and support all case studies. 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 aggregated nature. Underpinning such analyses is a hierarchy of ABS geographical classification boundaries to which statistical datasets are related. The geospatial interface to the AURIN portal allows users to drill into national, regional and local/suburban urban data sets. This interface exploits a local administrative geographic Gazetteer (a place-name lookup service), allowing directed adaptive zooming to a location of choice as displayed in Figure 1. The connection with an advanced gazetteer, providing access to other administratively named spatial entities, will be integrated in the following Lenses.


AURIN - liferay.com - Mozilla Firefox
File Edit View History Bookmarks Tools Help
AURIN - liferay.com
http://115.146.90.140:8080/
Welcome
AURIN
liferay.com
Open Layers Map
1000 km
500 mi
200 km
100 mi
2 mi

Location Selector

Pick the locations you're interested in...
Select a State: --select one--
Select a LGA: --select one--
Select a Suburb: --select one--

Dataset Selector

Get the feature type description for this capability:
--select one--

Location Selector

Pick the locations you're interested in...
Select a State: Victoria
Select a LGA: --select one--
Select a Suburb: --select one--

Dataset Selector

Get the feature type description for this capability:
--select one--

Location Selector

Pick the locations you're interested in...
Select a State: Victoria
Select a LGA: Melbourne (C)
Select a Suburb: --select one--

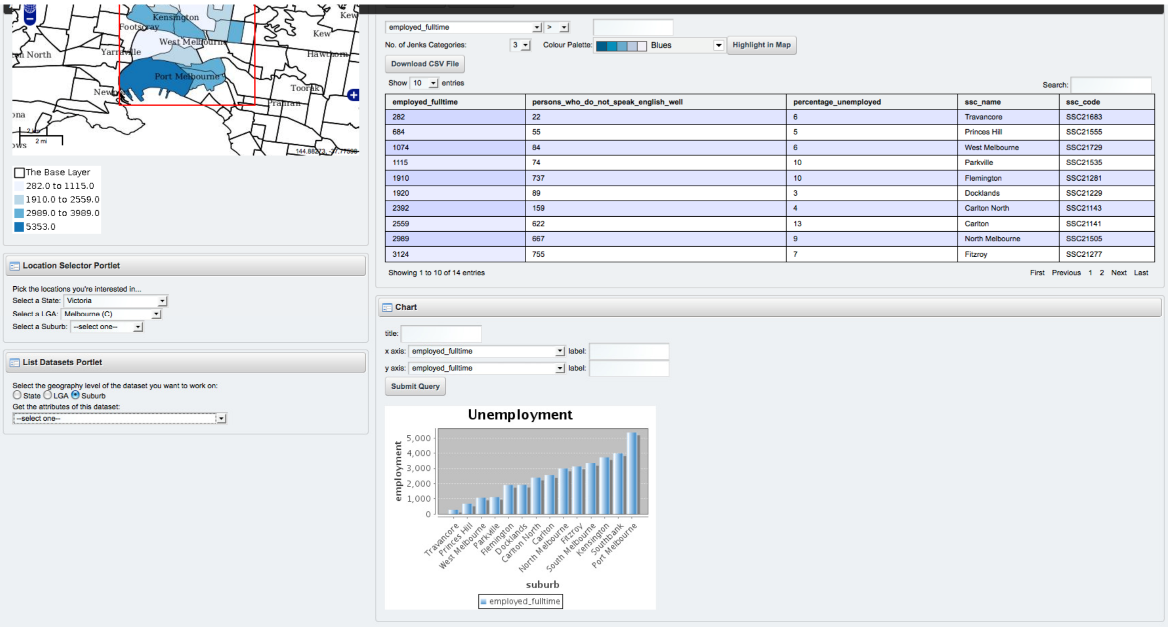
Dataset Selector

Get the feature type description for this capability:
--select one--

Figure 1. Drilling down hierarchically through the Australian geographical classification.

Currently available datasets comprise a variety of ABS aggregate level variables from the 2006 Census (Population, Income, Household and Employment statistics), and are accessed through a remote WFS service provided by the SLIP portal of Landgate WA (www.landgate.wa.gov.au). Additional local datasets include a sample from the Public Health Information Development Unit of the University of Adelaide (www.publichealth.gov.au/). These datasets can be used in typical scenarios, such as analyzing (see Figure 2):

- 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 Melbourne.



□ The Base Layer
282.0 to 1115.0
1910.0 to 2559.0
2989.0 to 3989.0
5353.0

Location Selector Portlet

Pick the locations you're interested in...
Select a State: Victoria
Select a LGA: Melbourne (C)
Select a Suburb: --select one--

List Datasets Portlet

Select the geography level of the dataset you want to work on:
○ State ○ LGA ● Suburb
Get the attributes of this dataset:
--select one--

employed_fulltime >
No. of Jenks Categories: 3 Colour Palette: Blues Highlight in Map
Download CSV File
Show 10 entries Search:

employed_fulltime	persons_who_do_not_speak_english_well	percentage_unemployed	ssc_name	ssc_code
282	22	6	Travancore	SSC21683
684	55	5	Princes Hill	SSC21555
1074	84	6	West Melbourne	SSC21729
1115	74	10	Parkville	SSC21535
1910	737	10	Flemington	SSC21281
1920	89	3	Docklands	SSC21229
2392	159	4	Carlton North	SSC21143
2559	622	13	Carlton	SSC21141
2989	667	9	North Melbourne	SSC21505
3124	755	7	Fitzroy	SSC21277

Showing 1 to 10 of 14 entries First Previous 1 2 Next Last

Chart

title:
x axis: employed_fulltime label:
y axis: employed_fulltime label:
Submit Query

Figure 2: Suburb Population vs Income around Melbourne

The phenomena studied can be displayed on a map including automated classification and colour-coding of mapped information using proper cartographic visualization codes, scientific graphing, and data inspection in a spreadsheet-like environment. Forthcoming features include regression analyses and enhanced graphing capabilities.

6. CONCLUSIONS

In this paper we have described the initial stages of the development of the AURIN e-Research Infrastructure. Despite being in the early days of development, the approach taken focuses on a highly pragmatic, demonstrator-based identification of sought-after functionality. As individual capabilities are added, the system will evolve into an infrastructure-enabling advanced analyses, simulation and modelling aimed at urban researchers.

The AURIN operates in the context of other major e-Infrastructure activities 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 engaging directly with these projects for the delivery of much of its underpinning infrastructure.

Forthcoming AURIN functionality is focused on enhancing the robustness of the platform and the integration of a workflow environment capable of integration with popular open source statistical analysis packages, such as the R project. Advanced graphing capabilities, data import and export features, an agent based modelling platform, 3D building and environmental modelling and visualization and similar functionalities are all scheduled for the following years. The AURIN project runs until mid 2014.

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

AURIN (2010). AURIN Final Project Plan, Available at: http://aurin.unimelb.edu.au/__data/assets/pdf_file/0005/463982/AURIN_Final_Project_Plan.pdf

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

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, de Roure, R. D., Sinnott, R.O. (2010). The Elements of a Computational Infrastructure for Social Simulation, *Journal of the Philosophical Transactions of the Royal Society A*, DOI:10.1098/rsta.2010.0150.

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

Correa-Velez, I., Gifford, S. M., and Bice, S. J. (2005). Australian health policy on access to medical care for refugees and asylum seekers, *Journal of Australia and New Zealand Health Policy* pp. 2 – 23, 2005.

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

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

Lambert, P., Tan, L., Turner, K., Gayle, V., Sinnott, R.O., Prandy, K. (2007). Data Curation Standards and Social Science Occupational Information Resources*, International Journal of Digital Curation*, Vol 2 (1).

Liao, E, Shyy, T. 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*, Vol 54 (2): 63-88.

Malleson, N., Birkin, M. (2011, forthcoming). Towards Victim-Oriented Crime Modelling in a Social Science e-Infrastructure. *Philosophical Transactions of the Royal Society A*.

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

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.

Purves, R. S., Edwardes, A., Fan, X., Hall, M., & Tomko, M. (2010). Automatically Generating Keywords for Georeferenced Images. In Proceedings of the GIS Research UK Conference, London, UK.

Sinnott, R.O., 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.

Schut, P., (2007). OpenGIS Web Processing Service. , (OGC 05-007r7). Available at: http://portal.opengeospatial.org/files/?artifact_id=24151.

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,* Issue 368: 3761-3778.