

Adapting Geospatial Business Intelligence for Regional Infrastructure Planning

Rohan Wickramasuriya^a, Pascal Perez^a, Jun Ma^a and Matthew Berryman^a

^a SMART Infrastructure Facility, University of Wollongong, Northfields Ave, Wollongong NSW 2522
Email: rohan@uow.edu.au

Abstract: Business Intelligence (BI) has traditionally been used in organizations as a strategic tool to maximize profit. When coupled with Geographic Information Systems, however, BI can be transformed into a cutting edge decision support system for planning local and regional areas, as we demonstrate in this paper. Local and regional governments often face a major challenge in terms of developing a holistic view upon disjointedly operated utility services in their jurisdictions due to data silos. This limitation has become a serious impediment to infrastructure planning and regional adaptation to changes. Geo-BI provides tools to manage data coming from multiple and disparate sources, and visualize them through online interactive user-interfaces. The SMART Infrastructure Dashboard (SID) is an innovative Geo-BI solution that includes an open-source ETL (Extract, Transform and Load) toolkit to handle various datasets, a spatially-enabled data warehouse hosted in PostgreSQL/PostGIS and proprietary BI software for creating and administering analytical reports and dashboards. SID allows planners and policy makers to analyze the interplay between the use of infrastructure services, demographics and weather parameters across multiple spatial and temporal scales. Furthermore, SID enables planners to run various what-if scenarios related to projected consumption patterns, service vulnerability of utility networks, and transportation demand management. Future research involves enabling the analysis of networks of networks through SID to understand the propagation of cascading failures and benefits in interconnected utility networks.

Keywords: *Geospatial Business Intelligence, Infrastructure, Regional Planning*

1. INTRODUCTION

Provision of infrastructure services to communities is a fundamental requirement, and this has been traditionally viewed as the responsibility of governments at various levels. However, in many developed countries, including Australia, the private sector is increasingly dominating the provision of such services (Cannadi and Dollery, 2005). While this increasing private sector involvement eases the pressure on local and state governments' limited resources, it inevitably brings new challenges in terms of monitoring and regulating services provided by several disjointed organizations. These challenges are further exacerbated by the fact that modern infrastructure networks are highly interconnected (Wang *et al.*, 2012). Hence, local and state governments urgently need an integrated view on infrastructure networks and services for better governance and planning of cities and regions.

Tools that aim to provide this integrated view should overcome several technical challenges. First, given the diversity of infrastructure networks and the decentralized nature of their management, they should be able to harness diverse and dispersed data sources. Second, they have to handle the underlying complexity of operations on individual networks, as well as the interconnectedness of networks. Moreover, as any other decision-support system, these tools need to exhibit positive usability traits such as performance, user-friendliness and intuitive visuals.

In this study, we demonstrate that Geospatial Business Intelligence (Geo-BI) can fulfill these requirements. Business Intelligence (BI) refers to “the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance” (Gartner, 2012). In terms of processes, a BI project involves data acquisition, data warehousing, data analysis and mining, and reporting and presentation (Shi *et al.*, 2012). Geo-BI is an improvement upon traditional BI made possible by integrating Geographic Information Systems (GIS) with BI (Angelaccio *et al.*, 2012). This integration, though technically challenging, opens up a myriad of new and exciting ways to analyse and present data.

Our approach to Geo-BI marks a radical change to the way BI has been traditionally used. Both BI and Geo-BI are conventionally used to deal with data from a single organization with the aim of improving profit and performance. The use of Geo-BI to access data from multiple providers and operators in order to analyse interdependencies of infrastructure systems and services is relatively recent (Sinnott *et al.*, 2011; Mokrech *et al.*, 2012). Further, the use of Geo-BI in this context is not only targeted at improving the profit and performance of individual infrastructure services, but enables policy makers and stakeholders to view the response of an integrated suite of interconnected utilities and services, to different scenarios, based on a complex-systems approach.

The aim of this study is to develop a robust, easily accessible and user-friendly Geo-BI solution, the SMART Infrastructure Dashboard (SID), which can harness diverse and dispersed datasets to support decision making related to local governance of infrastructure services. SID aims to inform planners and policy makers about the current and past states of infrastructure systems and services, as well as their spatial and temporal interdependencies. SID also enables future planning by allowing users to run various ‘what-if’ scenarios based on user-defined parameter values.

To develop SID, we first identified a geographic area of interest (the Illawarra region in New South Wales, Australia) that is typified by multiple infrastructure service providers. Then we convened a series of workshops with the participants from the infrastructure service providers and from the relevant councils to understand the available data and requirements for SID. Once the requirements were stipulated, we evaluated a number of existing and popular software frameworks to see if these can fulfill the stipulated requirements. It was clear that no individual existing framework could meet all the requirements, and we needed to bring several components together innovatively as discussed in the next sections. Once SID was put together, we developed several applications in SID following stakeholder consultation. Section 3 describes some of these applications based on which users evaluate SID. We adopt the pathway of continuous user consultation to evaluate SID followed by appropriate enhancements.

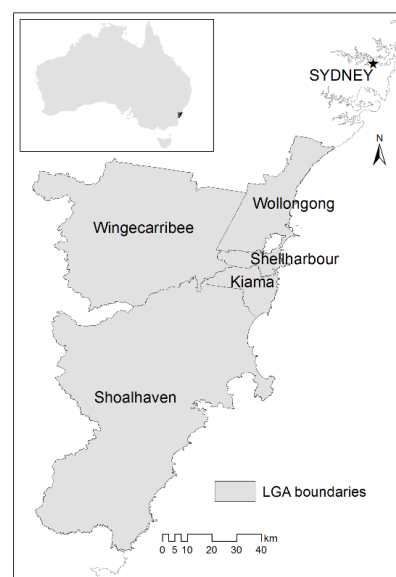


Figure 1. The Illawarra Regions, New South Wales, Australia

2. THE ILLAWARRA REGION

The Illawarra, a coastal region located south of Sydney, is made of five Local Government Areas (LGAs): Wollongong, Shellharbour, Kiama, Shoalhaven and Wingecarribee (Figure 1). The first four LGAs occupy the coastal plain limited on the east by a forested cliff, while Wingecarribee LGA spreads across the southern tableland, west of the cliff. According to the 2011 census (ABS, 2011), the population of the Illawarra region stood at 413,216 persons, 46.6% of which lived in Wollongong LGA only (Wilkinson, 2011).

Although the geography and topography of the Illawarra region have helped to create clear delineations for each utility network relatively well separated from neighbouring regions, authority and management vary considerably across utilities and jurisdictions. For example, the electricity distribution network is managed by a single operator (Endeavour Energy) for the whole region while water distribution is split between a private operator (Sydney Water, servicing Wollongong, Shellharbour and Kiama LGAs) and two local agencies (Wingecarribee and Shoalhaven LGAs). Likewise, a single private operator (REMONDIS) manages solid waste collection in Wollongong and Shellharbour LGAs while the three other LGAs administer their own facilities.

3. SMART INFRASTRUCTURE DASHBOARD

3.1. Stakeholder and Data Types

We identified the five LGAs and aforementioned private operators as stakeholders in the SID project. From SID's perspective, these stakeholders were both data providers and users. We collected a diverse set of data from them including geometric datasets of utility networks, service usage or consumption at various geographic levels over various time periods, water discharge at reservoirs and pumps, water quality at various points in the network, power consumption of assets such as treatment plants and pumps, waste collection routes, and quantity of waste collected. As early interactions with stakeholders showed their interest in correlating utility data with demographic and climate variables, we identified relevant databases from the Australian Bureau of Statistics (ABS) and the Bureau of Meteorology (BOM) to be incorporated into SID.

3.2. Technical Architecture and Work Flows

Figure 2 gives an overview of SID's technical architecture and main workflows involved. SID receives data from several providers in diverse file types (e.g. Excel spreadsheets, plain text files, CSV files, ESRI shapefiles) and in heterogeneous structures (e.g. number and types of columns).

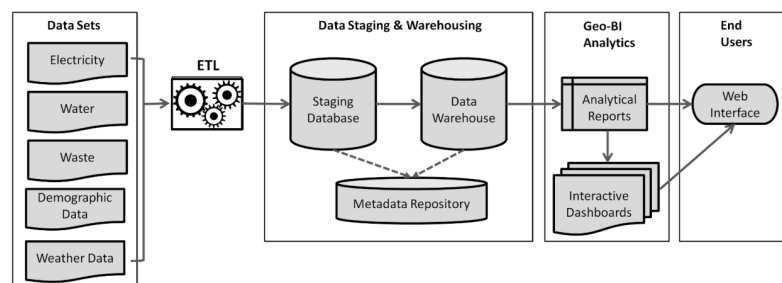


Figure 2. Technical Architecture and Workflows of SID

We use the standard Extract, Transform and Load (ETL) process (Kimball & Caserta, 2004) to migrate these data into an optimized data warehouse environment.

The ability to conduct analysis at multiple spatial scales has been identified as an essential component of SID. ETL plays a crucial role in shaping data to give SID this ability. Currently, two geographic hierarchies are supported in SID. The first is based on the Australian Statistical Geography Standard (ASGS) released by ABS recently (ABS, 2011). This hierarchy starts at Statistical Area Level 3 (SA3), and drills down to SA1 through SA2. The second geographic hierarchy is a two-step one where an LGA drills down to postcodes.

The data warehouse is based on a Star Schema design. Star Schema is a widely used data model for data warehousing in various real-world applications (Kimball and Caserta, 2004). It organises data into one or more fact tables referencing any number of dimension tables. A star schema can simplify join queries and provide capability to analyse data from multiple dimensions, thus enabling a user to perform various drill down, roll up, slice and dice operations on data. Figure 3 illustrates two linked star schema used in SID data warehouse. These two schema are linked using common dimension tables.

The dimension table ASGS models the geographic hierarchy adopted by ABS in its 2011 census. Note that this dimension table contains polygon geometry fields to store spatial extents of each census area. Dimension asgs supports drilling down and rolling up operations based on two geographic hierarchies described

previously. The dimension date models the date based on the ordinary calendar and the Australian season and financial calendar. This dimension enables multi temporal analysis of utility consumption. The fact table weather models temperature and rainfall derived from BOM daily records. Point source weather data collected from BOM were converted into surfaces using Kriging techniques prior to extracting average values at SA1 level. The fact demography models demographic information extracted from and generated based on ABS census results. The ABS conducts national census every five years. Based on the census results released in 2006 and 2011, demographic data in non-census years has been generated using linear interpolation.

In our data warehouse, each utility networks is modelled using a separate star schema. However, these schema and the schema for utility consumption are interconnected through common dimensions. For clarity, only the star schema for water network is depicted in detail in Figure 3. The schema for water supply network consists of 10 fact tables and 7 dimension tables. This schema allows the user to perform several analyses: tracing quality of supplied water from household level all the way upstream to the treatment plant, querying the pumping cost for a unit volume of water from the source to a household, and running what if scenarios such as finding all affected SA1 areas given the failure of a particular pump or a water main.

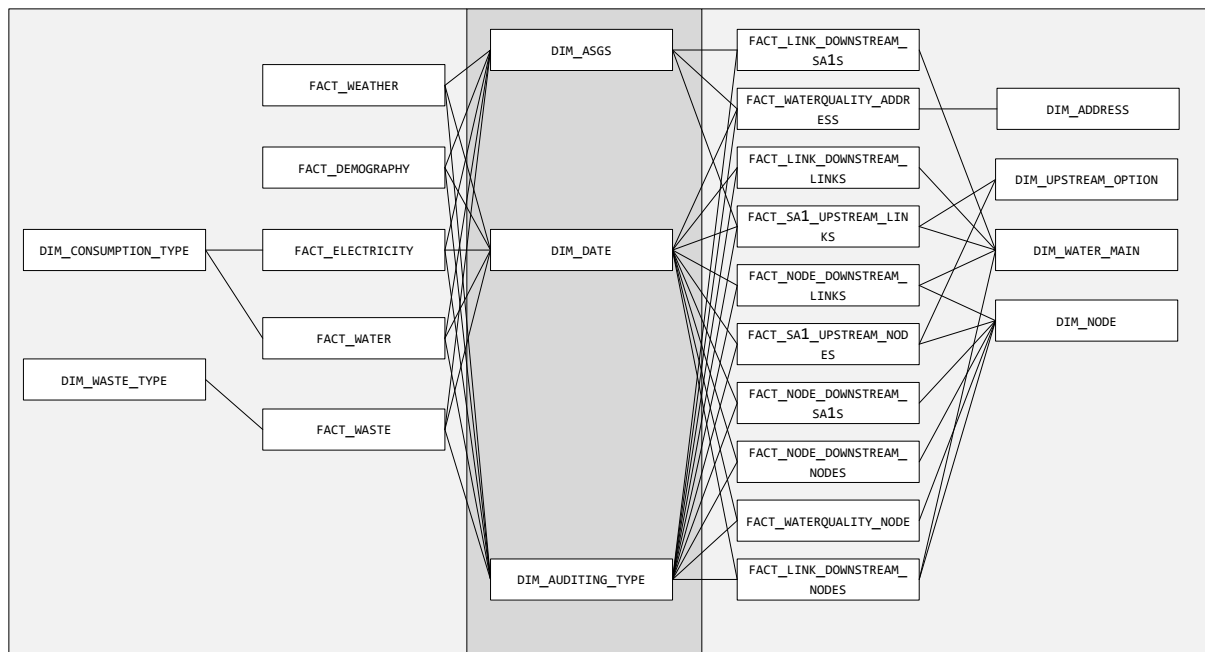


Figure 3. Overview of the star schema used in SID

Geo-analytical tools (proprietary Geo-BI software *Yellowfin* in SID's case) are used to create interactive reports such as maps and charts, and interactive dashboards that assemble related reports. These reports are built on top of the optimized star schema-based data warehouse. End users access this interactive content via an online portal. The user interface is visually rich and comprises of easy-to-use controls like filters that provide keys to intricate analysis while effectively concealing the complexity of calculations and database queries from the user. Visualization is a primary focus for the user interface as highly intuitive visuals play a pivotal role in successful policy support tools (Lindquist, 2011).

3.3. Applications

SID brings together disparate datasets on infrastructure service provision and use, demographics and weather allowing users to perform multi-dimensional, cross comparative analysis and visualization in a easy-to-use online platform. As a result, SID has a myriad of potential applications in the infrastructure domain that can provide planners and policy makers with ingredients needed to ensure a better service provision to communities. We briefly discuss two such applications next.

a) Relationships among Utility Usage, Socio-Demographics and Weather

Spatio-temporal pattern recognition and establishing association rules among variables are fundamental areas in spatio-temporal data mining (Andrienko *et al.*, 2006). Knowledge generated in these two areas has

practical implications for planning and policy making. For example, the knowledge that a certain group of suburbs consumes an increasingly high level of energy during winter can lead to several intervention actions. This could be an area-specific programme to retrofit energy efficient equipment or could be an investment in the network infrastructure to cope with the increasing demand. In order to generate such knowledge, analytical tools should support multi-spatial and multi-temporal analysis. Figure 4 shows an analytical dashboard developed in SID that caters for this demand.

This dashboard consists of three interactive reports. The map report starts at LGA level, and is drillable to postcode level. Moreover, the other two reports can be controlled by this map report. A user can select one out of several metrics, for example total residential electricity consumption, to be displayed on the map. The second report (radar chart) gives an overview of the utility usage for the active area on the map report. The statistical bubble chart shows the relationship between per capita water and electricity consumptions at postcode level. A bubble in this report represents a postcode, and the radius of a bubble is proportional to the population in that postcode while the colour of bubbles indicates the abundance of flats and units in a postcode. Filters to the right of the dashboard let users control all reports at once.

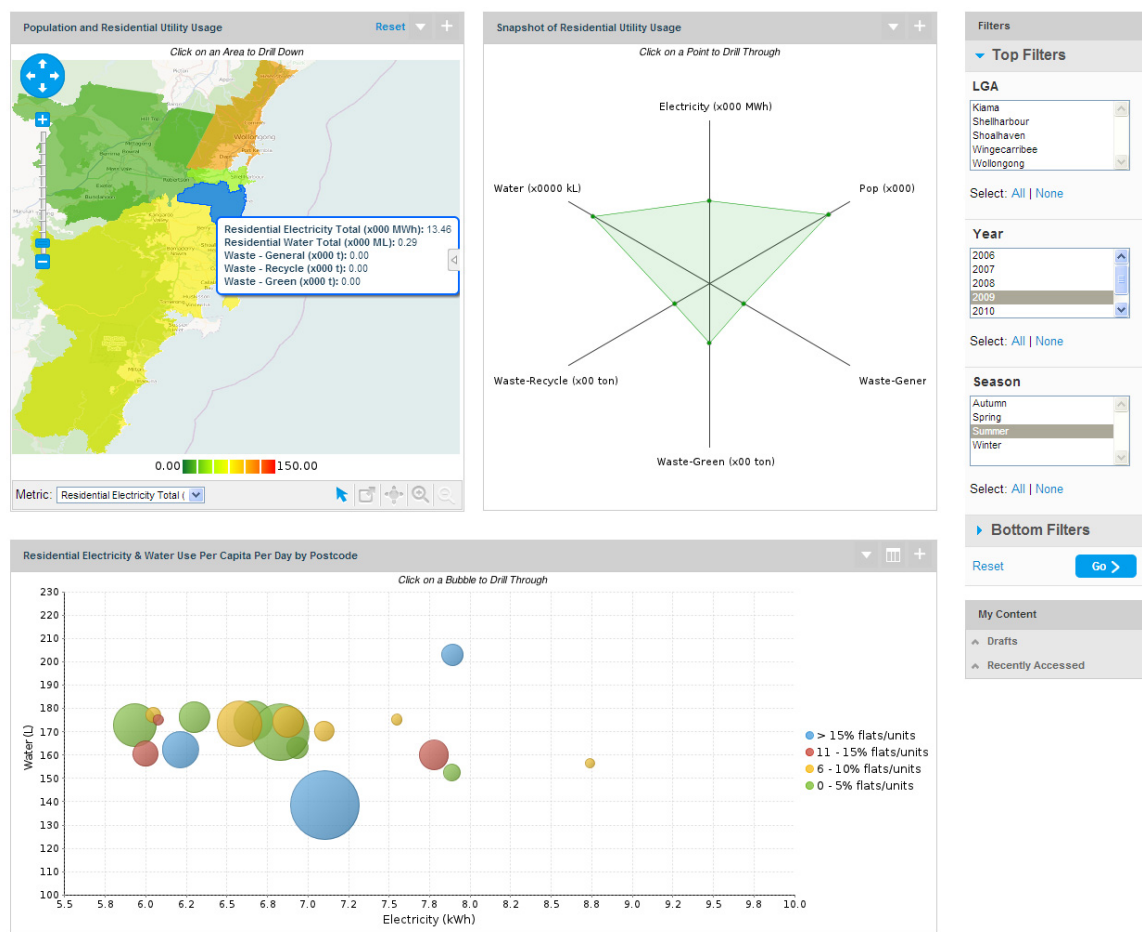


Figure 4. An interactive, geo-analytical dashboard used in SID

Multi scale spatial analysis is facilitated by the drillable map report, and the ability to update linked reports with the drill down or roll up on the map. Figure 5 illustrates how a user could click on a map polygon to visualize information at an increasingly finer granule. This has been made possible by the spatial hierarchy designed at the star schema-based data warehouse.

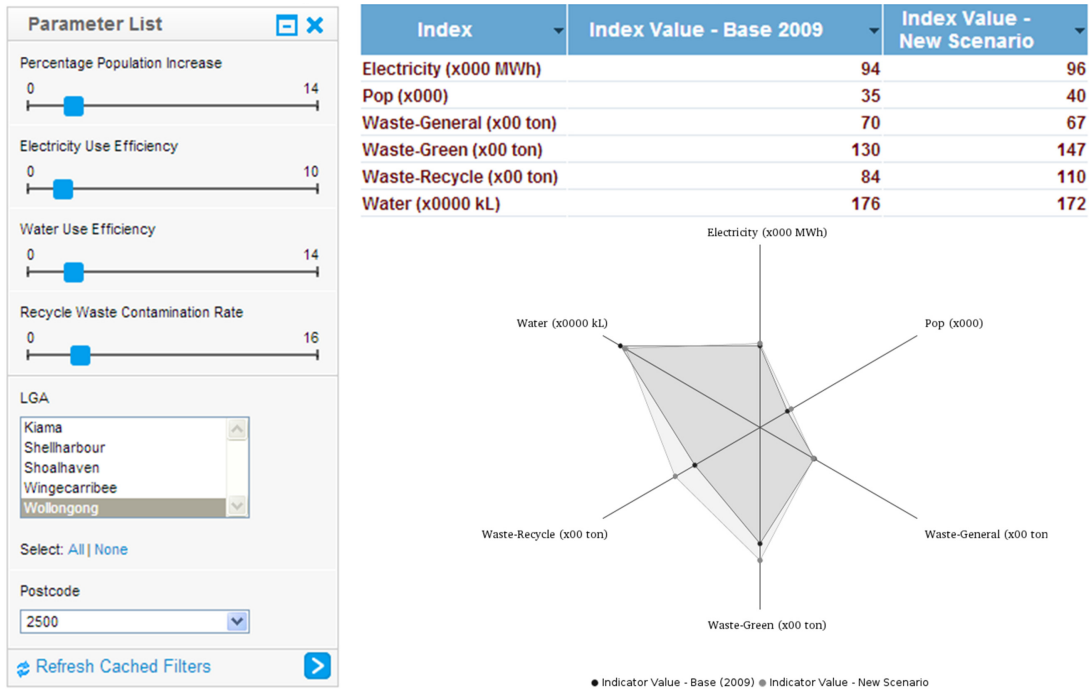


Figure 7. What-If Scenario Analysis

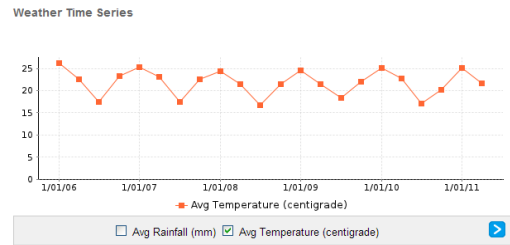
Multi temporal analysis is enabled in this dashboard through a hierarchical filter. For the temporal analysis, two levels of disaggregation are possible: year and the seasons. Although SID facilitates further disaggregation in the time dimension through month, weeks, days up to minutes, this particular application limits this disaggregation to season level only. The main reason for this was the water consumption data for which the finest available temporal granule was season.

Perhaps, the most important feature of SID is its ability to cross spatial and temporal granules at will. For example, some patterns could be pertinent at postcode level only when viewed at seasonal intervals, while some other patterns may project well at LGA level at yearly intervals. SID has been designed to uncover such hidden patterns in data.

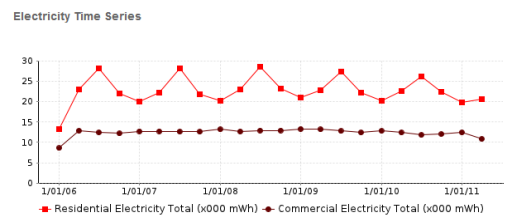
Statistical bubble chart in Figure 4 is an example for several ways by which SID uncovers interrelationships among various utilities and demographic variables at multiple temporal resolutions. While this chart enables the user to appreciate general trends, it is also an ideal tool to spot outliers. SID then supports the natural curiosity of the analyst to find more about this outlier via detailed 'drill through' reports (Figure 6).

b) *What-If Scenario*

What-If analysis is an indispensable tool by which planners could estimate the potential impact of changes in a set of independent variables on one or more dependent variables (Golfarelli and Rizzi, 2009). Traditionally, BI solutions have been designed to analyse past data only, and they lack the ability to provide any sort of anticipation of future trends. Hence, what if analysis is a modern and cutting edge feature of BI that makes such tools sit on the top of the BI pyramid (Golfarelli and Rizzi, 2009). SID as a Geo-BI solution provides planners with ample opportunities to leverage historical data for better planning through what if analysis. Figure 7 shows one of several integrated reports in SID that let users to carry out what if analysis.



(a)



(b)

Figure 6. Child reports illustrating the relationship between utility use and weather at postcode level, (a) Weather time series, (b) Electricity time series

Using this integrated report, a planner can estimate the expected utility use for a given LGA or postcode under various scenarios, and compare the predicted value with the base case both tabularly and graphically. While the

planner predicts the utility usage in 5 years, for example using a certain value for the population increase, he could also see what efficiency gains would keep the expected usage within manageable limits. This report keeps complexity of the implemented calculations behind the scene, and presents the non-savvy user with a few easy-to-use sliders and filters to perform multiple scenario analyses on utility usage.

4. DISCUSSION AND CONCLUSIONS

Infrastructure services at the local and state levels are usually provided by a mix of public and private agencies. Private sector is increasingly dominant in this domain particularly in developed countries like Australia. Nonetheless, it is the responsibility of local and state governments to ensure that the public receive the best possible infrastructure services. These public agencies, therefore, urgently need an integrated view on the provision and use of infrastructure services in their legislative areas. Among several reasons for the lack of such integrated view, the difficulty to collate disparate datasets from multiple agencies and the technical challenges associated with developing such solutions due to intricate network structures and operations of interconnected infrastructure systems are at the pinnacle.

We demonstrate that the Geospatial Business Intelligence can be adopted to provide the much needed integrated view on infrastructure service. Geospatial aspect is crucial in this solution to take the best out of majority of the infrastructure-related data that contain a location reference. Moreover, identifying patterns and associations among various infrastructure services and usage in space-time is crucial in decision making. Needless to mention that an interactive map is a better visual when it comes to communicating location-related information. BI provides tools and methods needed to tap into diverse sets of disparate data and load them into an optimized data warehousing environment for efficient analysis and reporting. With the involvement of right people from the design phase to final usage phase, Geo-BI can be turned into a powerful tool for the governance of infrastructure services.

Using a case study for the Illawarra region of New South Wales, Australia, we demonstrated how tools and processes in Geo-BI could be harnessed to develop a user-friendly solution, which we call the SMART Infrastructure Dashboard (SID), geared towards the governance of infrastructure services at the local and regional levels. Through a web-based portal, SID provides planners and policy makers a visually-rich interface to perform powerful spatio-temporal analyses needed to identify patterns and associations among multiple utility-related variables in space and time. Moreover, SID facilitates what if scenario analysis offering a way by which planners could anticipate and plan for future trends in utility usage. Future research will involve extending SID's capabilities to investigate the propagation of cascading failures in interconnected networks.

ACKNOWLEDGEMENT

Authors would like to acknowledge Sydney Water, Endeavour Energy, REMONDIS and the Southern Councils Group. This research has been supported by the Australian National Data Service (ANDS) through their Application program.

REFERENCES

- ABS, 2011. Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas. Australian Bureau of Statistics, Canberra.
- Andrienko, G., Malerba, D., May, M., & Teisseire, M. 2006. Mining spatio-temporal data. *Journal of Intelligent Information Systems*, 27, 187-190.
- Angelaccio, M., Basili, A., Buttarazzi, B., & Liguori, W., 2012. Using geo-business intelligence to improve quality of life, IEEE 1st AEES European Conference on Satellite Telecommunications. 02 - 05 October 2012, Rome.
- Cannadi, J., & Dollery, B. 2005. An evaluation of private sector provision of public infrastructure in Australian local government. *Australian Journal of Public Administration*, 64, 112-118.
- Gartner, 2012. IT Glossary. Gartner Publishing, Stamford.
- Golfarelli, M., & Rizzi, S. 2009. What-if simulation modeling in business intelligence. *International Journal of Data Warehousing and Mining*, 5, 24-43.
- Kimball, R., & Caserta, J. 2004. *The Data Warehouse ETL Toolkit*. Wiley Publishing, Inc.
- Lindquist, E., 2011. Grappling with Complex Policy Challenges: exploring the potential of visualization for analysis, advising and engagement.
- Mokrech, M., Nicholls, R. & Dawson, R. J. 2012. Scenarios of future built environment for coastal risk assessment of climate change using a GIS-based multicriteria analysis. *Environment and Planning B: Planning and Design*, 39 (1): 120-136.
- Shi, H., Peng, C., & Xu, M. 2012. Business intelligence in construction: A review. *Advanced Materials Research*, 594-597, 3049-3057.
- Sinnott, R. O., Galang, G., Tomko, M. & Stimson R. 2011. Towards an e-Infrastructure for Urban Research across Australia. IEEE e-Science Conference, Stockholm, Sweden. Access online: <http://aurin.org.au/wp-content/uploads/2011/10/AURIN-v3.pdf>
- Wang, S.L., Hong, L., & Chen, X.G. 2012. Vulnerability analysis of interdependent infrastructure systems: A methodological framework. *Physica a-Statistical Mechanics and Its Applications*, 391, 3323-3335.
- Wilkinson, J., 2011. The Illawarra: An Economic Profile, in: N. Parliament (Ed.). Sydney.