Building a Web Observatory for South Australian Government: Supporting an Age Friendly Population

Xin Wang[1], Ramine Tinati[1], Wolfgang Mayer[2], Anni Rowland-Campbell[3] Thanassis Tiropanis[1], Ian Brown[1], Wendy Hall[1], Kieron O'Hara[1], Markus Stumptner[2], and Andy Koronios[2]

[1] Web and Internet Science Group
Electronics and Computer Science
University of Southampton, UK
http://www.ecs.soton.ac.uk/
{xin.wang,R.Tinati,tt2,ian.brown,wh,kmo}@ecs.soton.ac.uk
[2] School of Information Technology and Mathematical Sciences
University of South Australia
http://www.unisa.edu.au/it-engineering-and-the-environment/information-technology-and-mathematical-sciences/
{Wolfgang.Mayer,Markus.Stumptner,Andy.Koronios}@unisa.edu.au
[3] Intersticia
http://intersticia.com.au/
anni@intersticia.com

Abstract. The South Australian Government has played a leading role in the digital government trend, and Adelaide has led the move towards becoming a "smart city" ever since the "multi-function polis" in the early 1990s. As a part of the move towards Open Government many jurisdictions have created platforms upon which to publish open public data, and South Australia's data.sa.gov.au is where its public government data are made available as csv files for use by business and the community. However, there is much more to opening up data than just publishing it, because when it comes to data that contains sensitive and private information, the current data publishing and sharing solutions hit barriers due to the lack of data provenance and security mechanisms, as well as limited use/usability for applications. Further, building analytic applications that make use of those datasets in line with their sharing policies is another challenge as many users have little digital literacy. These issues can be addressed by a new data sharing platform called the Web Observatory which allows both data and analytics to be published, shared and consumed in a secure and transparent manner. This paper describes the joint effort of building a Web Observatory for the South Australian Government through the collaboration of the University of South Australia and the University of Southampton.

Keywords: web observatory, semantic catalogue, digital government

1 Introduction

The World Wide Web is arguably one of the largest, and richest resources for obtaining and generating data, information, and human activity. It's global reach, and diversity in use and content, has led to it becoming a crucial information asset for connecting and supporting the growing and changing population of the World.

Originally deployed via a Web browser, the document-driven Web has evolved into a ubiquitous technology that is part of a growing eco-system of devices, protocols, and human activity. Despite being only 25 years old, this new ecosystem of interactions and data now offers the potential to transform how we, as individuals, and as a collective, interact. Moreover, we have already started to see the benefits of a connected, Web-enabled society, especially with the turn towards the publication of Open (Government) Data, an initiative where governments - as well as businesses and individuals - are publishing their data in open, machine readable, re-usable formats. Open Data, and in particular Open Government Data, opens up a new avenue for transforming society, improving government-to-citizen transparency, and importantly, improving the lives of citizens everywhere.

As the demands on government change to accommodate changing societal demographics and needs the imperative for evidence-based service design informed by evidence-based policy is being recognised as an emerging priority. Open data platforms can partly address elements such as getting non-sensitive data *to* the population but not all data sets, particularly those around medical care and other personal data, are inherently open, nor do these platform address the collection or amalgamation of data *from* the population or from other sources. Systems which combine open/closed data sets and support a greater level of access control are required to meet the twin requirements of gathering/combining multiple data sources with that of protecting personal privacy.

In recent years, one of the key issues that many societies and countries have faced is the transition towards an ageing population, a transition in society where the proportion of older citizens are accompanied by a reduction in the proportion of children and the working ages [11]. By 2045 the number of older persons is expected to exceed the number of children globally for the first time. This crossover point was already reached regionally in 1998 in the more developed regions of the world where population ageing is more advanced.

The challenges presented by an ageing population have been extensively studied and documented at a global scale for more than 50 years including a landmark report by the United Nations [11] looking at population from 1950-2050 and a series of interim updates (most recently in 2009). These report the same ageing phenomena which are described as unprecedented, pervasive (occurring across almost all countries), profound in terms of its effects socially, economically, financially and politically and, critically, persistent in that the trend has remained upwards since 1950s and shows no signs of reversing in the 2050 time frame. Other local country-by-country studies including those in Australia [1,4] have confirmed similar findings.

Whilst not all the effects of an ageing population are thought to be negative [2], Governments globally are faced with a long term planning challenge in which substantial changes to the age/sex demographic and needs/abilities profile of their constituents will be accompanied by a projected fall in the PSR (potential support ratio) which captures the number of working age (16-64) individuals for each post-working age (¿64). This figure fell globally from 12 in 1950 to 9 in 2009 and is projected to drop to 4 by 2050. In Australia this support ratio is projected [ibid] to fall to just 2.5 by 2041. In the aftermath of the financial crisis we have already seen policies on retirement age, pension policies and health care starting to change.

In this paper we report on the ongoing Web Observatory activities between The University of South Australia, The University of Southampton, and the South Australian Government together with aged care providers. The paper describes the implementation of a Web Observatory, the challenges, and the scoping and development workshops that were run in order to begin addressing the challenges of supporting a growing elderly population using the Web Observatory to inform both the technological infrastructure and sharing principles.

2 Issues and Challenges of Releasing Government Data

In recent years, governments around the world, including Australia, have begun to release selected data sets to the public through the Web. The Australian Commonwealth led the way in it's "Declaration of Open Government" in 2008[4] and the establishment of its Gov 2.0 Taskforce (http://www.gov2.net.au/) and State governments followed. In 2013, the South Australian government announced the "Open Data Declaration"[5] with the aim of making data openly available to the public by default, and supporting guidelines have been developed to assist agencies in identifying, describing, and releasing suitable data sets. Public events, such as "GovHack", aim to foster the use of open data for public and commercial purposes. These are laudable goals, but the reality is that the processes, tools, and utilisation strategies related to open data are not yet well understood, particularly beyond the technical community. In addition, legal and policy concerns with regard to what data to release, licencing and terms of use, ensuring data privacy, measuring success and impact, together with technical considerations regarding the format and method of publication remain to be addressed, and thus the willingness to experiment beyond clearly understood domains remains limited. In the following we focus on selected technical considerations which could provide a solution and therefore a framework for greater publication, sharing and utilisation of data by both researchers and public policy practitioners to achieve these goals.

[4] http://www.finance.gov.au/e-government/strategy-and-governance/gov2/declaration-of-open-government.html

[5] http://dpc.sa.gov.au/open-data

Numerous data repositories have been established that support the open data initiatives across Australia.[6] Foremost, the Australian government's federated data hub data.gov.au makes available a large number of data sets generated by government agencies at state and federal levels. The Australian National Data Service (ANDS) maintains a broad catalogue of scientific research data sets that have been generated by Australian research institutions. The Research Data Storage Initiative (RDSI) provides data storage facilities for research groups. Although federated repositories allow one to locate and retrieve data sets, making effective use of data sets remains challenging.

Making effective use of data sets remains challenging due to variability in data presentation. Most repositories operate as stand-alone sites with proprietary access mechanism and data formats. Data sets are released in a variety of electronic formats, at varying level of granularity, organised in arbitrary manner, and subject to a breadth of licencing conditions. The absence of sufficient meta-data about how data sets were obtained, for what purpose, and potential limitations in the data may pose difficulties when assessing whether data sets are suitable for a particular purpose.

Data access mechanisms further limit linking and analysis of data sets. Whereas most data hubs readily export data in a variety of formats (CSV, MS Excel, GeoJSON, and KML are commonplace), few support direct programmatic access to the data. Having to export data imposes further hurdles to making effective use of data, as effort must be expended to replicate the data in a form that can be queried, linked with other data sets, and visualised. Data quality issues may arise if the underlying data source changes, and any changes in data formatting may disrupt sustained replication of changing data.

Integration of analytic tools and published data sets remains an important area for further work. Powerful portals offering analytic capabilities and programmatic access to consolidated data hubs for specific areas of research have been developed under the stewardship of dedicated data custodians and system developers. For example, the Australian Urban Research Infrastructure Network (AURIN) network offers uniform access to a breadth of data sources related to urban development and integrated geo-spatial analytic tools to support scientific workflows and discovery [7]. As such AURIN has the potential of becoming an important resource as part of a wider Web Observatory network; however, further work would be needed to make AURIN accessible to a wider audience and integrate its underlying data hub with analytic applications not hosted on the AURIN portal itself.

The majority of data released through Government data hubs are static snapshots of historic data pertaining to public infrastructure, political organisations, public offices, census data, population health, and public transport. Little "live" data is available that could be queried automatically. Although some agencies have begun to offer public programmatic interfaces to their databases, most

[6] A list of well known open data initiatives is available at http://www.finance.gov.au/blog/2013/10/26/government-data-landscape-australia/

come with proprietary interfaces and restrictive terms of use that prohibit the integration with automated applications hosted in Web Observatories.

In general, licencing restrictions, security and privacy concerns, and costs associated with operating such interfaces will need to be addressed before live programmatic access will become part of day-to-day data publishing policies in government. Currently, most live data sources are restricted to manual querying and primitive visualisation of data. Similar considerations apply to commercial data sources, such as social media, news, and streaming data sources. Web Observatories and related technologies for fine-grained access control and cloud-based infrastructure for data replication, curation, and analytic applications may be a key ingredient in the path towards achieving this vision. Development of suitable policies, KPIs and incentives for data providers remain important issues for governments and the private sector.

Further enhancements to current data hubs are needed if the use and impact of released data sets is to be assessed. The current data publication mechanisms are largely one way from a data provider towards a community of users. Once data is published, current platforms make it difficult to control and track who uses data sets and for what purpose. In addition to fine grained flexible access control, it would be desirable to capture research findings and applications that result from the published data sets. To date, few direct mechanisms exists for users and researchers who derive insights from that data to provide feedback to share their results with the data providers and indeed the wider community. This ultimately limits the ability of data providers to assess the impact their data has had on the wider community.

As such, the vision of a community-based network of Web Observatories in which data, analytics, and insights are all linked and shared remains an important guiding framework for policy makers, researchers, and systems developers alike. Along the way, important issues relating to data curation, systems interoperability, data- and meta-data modelling, security and privacy, policy making and measurement will all need to be addressed.

3 Open Government Data with Enhanced Security

Making data open to the public improves discoverability and reusability. Further it enables isolated data to be linked to form a network of rapidly increasing value (as suggested by Metcalfe's law [5]). As a specific case, open government data inherit all the above benefits, and in addition creates value in areas such as police impact measurement, government transparency etc. Although the more open data the merrier, some data, especially in the case of government data, are too sensitive to be accessed by everyone. Without appropriate access control, even data that can be open to part of the public, have to stay closed.

Most data of South Australia Government are shared as data files (e.g. CSV, MS Excel). It is straightforward to place access control on top of file sharing, however, as stated in last section, data consumers would benefit more from data in shared databases, which can be relational databases, NoSQL databases, graph

databases etc. These databases various on capabilities and interfaces of access control. The Web Observatory is able to abstract all the complexity away and provide a unified interface to end users.

3.1 Access Control Options

For each shared resource the access control put constraints on two aspects: the metadata, which refers to the record of the resource in a web observatory, and the actual contents of this resource. Constraints on metadata enforce that the resource is only visible to authorised users; constraints on content make sure only authorised users can access the resource itself. For example, a user who only has the permission to access to the metadata of a dataset can search this dataset via the web observatory, but cannot query this dataset.

A resource falls in one of the three access control categories: 1) *open*, that both metadata and contents of the resource are open to everyone; 2) *private*, that metadata are open to everyone, but contents are only open to authorised users, and 3) *do not list*, that users need authorisation to access both metadata and contents. A summary of the three options is given in Figure 1. By default the owner of resources always have access to both metadata and contents.

	Metadata	**Content**
Open	Everyone	Everyone
Private	Everyone	Authorised
Do not list	Authorised	Authorised

Table 1. Web Observatory access control options

The access control on metadata affects discoverability of resources. Web Observatory provides resource metadata as Schema.org micro data that can be recognised by compatible search engines (e.g. Google, Yahoo!) and applications. Those search engines and applications are regarded as unauthorised users, i.e. only open resources can be discovered.

4 Use Cases of Government Web Observatory

A Government Web Observatory offers many different services, internally for organisational benefits and cost-savings, and externally, civil transparency, the deployment of public services, and in certain scenarios, bridging the gap between government-to-citizen communication and engagement. If we consider the case of the South Australian Web Observatory, the ability for government to publish their data alongside other forms of datasets, whether this be Web data, social

media data, or device-driven data such as traffic and air pollution data, is a powerful capability.

Linking back to the earlier discussion regarding the societal challenges of dealing and reacting to an ageing population, the Web Observatory can be a powerful socio-technical infrastructure for government, providing the technical capabilities to answer questions about how best to support the changing demographics of a region's population. Making use of working within a distributed paradigm in which datasets are managed via various levels of access and control, complex questions can be asked across multiple data sources in order to ask questions and make decisions on how best to deal with this societal change.

As a recent report on caring for the elderly suggests [3], there are various ways in which government can measure the 'age-friendliness' of a city, and in some cases, develiver intervention mechanisms in order to support this growing demographic. There is also an increasing recognition that data-driven decision making is an important transition in policy planning.

In several of the recommendations discussed in this and other advisory reports [6] on the changing population demographics, there is a series of discussions regarding the use of data, both public and private in order to make informed and accurate decisions to improve the public services available. More over, data-driven decisions can inform internal government plans and future strategies. As these reports describe, answering questions about how best to cope and support an ageing population requires a spectrum of resources, including a balance between government published data, open data, and potentially private data, which may come from the private sector, or more generally, from services (on the Web) which are associated with various terms of conditions and contracts.

4.1 Working towards An Age-Friendly City

As part of an active scoping and development exercise between Government, and researchers at the University of Southampton and South Australia, we ran a workshop to consider how to use the Web Observatory, and the datasets available in order to answer questions about Adelaide as an age-friendly city. The attendees of the workshop included government experts working in digital transformation, as well as those directly involved in supporting communities within the city.

The workshop was guided by two example questions which were distilled from the Kalache report's recommendation, which were concerning (1) providing adequate leisure facilities and fitness equipment for the ageing population, and (2) to monitor, protect and improve the health and well-being of older citizens, and to ensure that they are socially connected within the community. Both these recommendations are part of a broader framework of questioning what public services are available within Adelaide, and how these can best be used to support the well-being of citizens and their connection within the community.

The process of answering the questions raised above involved working through the available datasets, asking questions about what *internal* and *external* value could be obtained by not only linking and analysing the data, and considering

the short and long term implications. In the first example (1), the availability of public facilities and services data proved adequate to roughly answer questions about how well equipped a district in Adelaide is for supporting fitness and leisure activities for elderly citizens. The focus group discussed the internal and external *products* that could results from combining several of the datasets listed in Table 2. For internal value, the main discussion involved using a visualisation and analysis of the location of public facilities (i.e. the Walking Trails and Bike Hire locations) in combination with the district demographic information in order to make informed decisions about the development of further services based on the average age of the district. For external tools, Web-enabled visualisations could be made to map the location of public facilities in order to provide residents of specific district within Adelaide with an overview of what services are available.

A number of scenarios were worked through, from providing pre-compiled biking and walking routes, based on the location of bike hire locations, transport stops, and rest stops. Although these are simple scenarios, these were considered as resources which are extremely useful for citizens who are wanting to take part in some leisurely activity but may not be aware of how, or where this can be conducted. For the internal and external outputs, the Web Observatory's infrastructure, specifically the ability to query multiple distributed datasets, and list and use public and private data was considered a key asset. In the example scenarios identified, the datasets (as listed in Table 2 required were all publicly available (via data.gov.sa) but not queryable. By listing them on the Web Observatory, the data could be queried, programmatically via the API or through the Web portal.

For example (2), the focus group identified a number of challenges in which the Web Observatory infrastructure offers the necessary features in order to answer specific questions, however, data availability was identified as a issue. In order to answer questions about an individuals health, well-being, and their social connectivity, the focus group derived a number of confidential datasets would be required in order to answer such questions. As show in Table 2, although there are several datasets regarding location-based demographics and hospital location data, to consider the health of the citizens at a granular-level, an individual's medical, and if possible, movement data would be required. Medical records and transport card data were identified as appropriate sources of data to support this, and using the Web Observatory sharing mechanism would mitigate the problems of confidential data. The access control layer would makes it possible to provide restricted control to these datasets in order to retain confidentiality and ownership. Social data from social networking platforms such as Twitter and Facebook was also considered as a valuable source of data to support this recommendation. Monitoring, communications and providing support via these systems offers another mechanism to improve the engagement of citizens within the community

As an internal government tool, the ability to examine - at an aggregate level - the well-being of a group of citizens based on their activity level and current

state of health becomes an extremely useful mechanism to pro-actively redirect resources to a given area when required. It also becomes a tool that can advise health-care workers on certain individuals who many require more attention due to their activity level and state of health.

Dataset	**Rec. 1**	**Rec. 2**
Local Hospital network		
Remoteness Area		
Demographic Local Area		
Park Land Sporting Facilities		
Parkland Parks		
Adelaide Metro Real-Time Passenger Information		
Seats		
Park Land Toilets		
Picnic Tables		
Monuments		
Bike and Pedestrian Paths		
Bike Hire Locations		
Adelaide Public Transport Stop Data		
Public Transport Services		

Table 2. Adelaide Open Data an Additional Datasets and their use in Supporting Age-Friendly Recommendations

5 Conclusions and Future Plan

The vision of a Web Observatory as a global, distributed resource [10,8] that will empower government, business and individuals with access to data and software resources for analytics will need to be shaped by incubation projects with different stakeholders. The South Australian Government Web Observatory is a significant project exploring ways in which government can build on the power of open data portals with (i) access to datasets that can be shared by individuals or organisations (ii) access to datasets that can be queried on datastores of a variety of datastores (iii) application programming interfaces (API) that enable the development of analytic applications using one or more public or shared datasets (iv) the means to access other datasets or analytic applications on the emergent Web of Observatories in a harmonised way [9].

This paper has focused primarily on the technical challenges of building a Web Observatory for government focusing on the particular needs of South Australian government. As we have seen with the invention of the Web, such small technological interventions when adopted (and inevitably shaped and evolved) on a large scale they can take digital literacy on a new level empowering citizens to engage with government in more meaningful and effective ways. They can also create value and markets previously unforeseen benefiting existing sectors or fostering the creation of new ones based on data and analytics for everyone.

The future steps are on providing key analytic applications that will answer key questions around demographics and the elderly as per the Kalache report

making the most of the datasets that are already hosted in the Web Observatory as well as those that are listed in other nodes of the Web Observatory network[7]. We expect those key applications to serve as catalysts for the engagement of government, organisations business and the public.

References

1. Commission, A.: An ageing australia. Tech. rep., Australian Government Productivity Commission (2013)
2. Healy, J.: The benefits of an ageing population. Australia Institute (2004)
3. Kalache, A.: The longevity revolution: Creating a society for all ages. Tech. rep., Government of South Australia (May 2013)
4. Karmel, T., Ong, K.: Will we run out of young men? implications of the ageing of the population for the trades in australia. In: Maclean, R., Wilson, D. (eds.) International Handbook of Education for the Changing World of Work, pp. 2445–2456. Springer Netherlands (2009), http://dx.doi.org/10.1007/978-1-4020-5281-1_161
5. Metcalfe, B.: Metcalfe's law after 40 years of ethernet. Computer 46(12), 26–31 (2013)
6. Select Committee on Public Service and Demographic Change: Ready for Ageing? HOUSE OF LORDS (2013)
7. Sinnott, R.O., Bayliss, C., Bromage, A.J., Galang, G., Grazioli, G., Greenwood, P., Macauley, A., Morandini, L., Nogoorani, G., Nino-Ruiz, M., Tomko, M., Pettit, C.J., Sarwar, M.S., Stimson, R., Voorsluys, W., Widjaja, I.: The australia urban research gateway. Concurrency and Computation: Practice and Experience 27(2), 358–375 (2015), http://dx.doi.org/10.1002/cpe.3282
8. Tiropanis, T., Hall, W., Shadbolt, N., de Roure, D., Contractor, N., Hendler, J.: The Web Science Observatory. IEEE Intelligent Systems 28(2), 100–104 (2013)
9. Tiropanis, T., Wang, X., Tinati, R., Hall, W.: Building a Connected Web Observatory: Architecture and Challenges. eprints.soton.ac.uk (2014)
10. Tiropanis, T., Hall, W., Hendler, J.A., de LarrinagaChristian: The Web Observatory: A Middle Layer for Broad Data. doi:10.1089/big.2014.0035 2(3), 129–133 (Sep 2014)
11. United Nations Department of Economic and Social Affairs, P.D.: World Population Ageing 2009. Tech. rep., Department of Economic and Social Affairs Population Division (2010), http://www.un.org/esa/population/publications/WPA2009/WPA2009-report.pdf

[7] http://webscience.org/web-observatory/list-of-web-observatories/