Computers, Environment and Urban Systems xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

An online platform for conducting spatial-statistical analyses of national census data across Australia

Christopher J. Pettit [a], Robert Tanton [b], Jane Hunter [c]

[a] City Futures Research Centre, Built Environment, UNSW Australia, UNSW Sydney, NSW 2052, Australia
[b] National Centre for Social and Economic Modelling, University of Canberra, University South Drive, Canberra, ACT 2601, Australia
[c] School of ITEE, The University of Queensland, Queensland, Australia

ARTICLE INFO

Article history:
Received 1 November 2015
Received in revised form 30 May 2016
Accepted 30 May 2016
Available online xxxx

Keywords:
Online platform
Open data
GIS
Spatial-statistical analyses

ABSTRACT

In recent years there have been significant advances in online planning and decision support systems that encompass spatial and statistical analysis and visualisation services. In this paper we provide an analysis and evaluation of one particular online spatial-statistical platform, known as the Australian Urban Research Infrastructure Network (AURIN) portal. By adopting the Statistical Data and Metadata Exchange (SMDX) format, more than 500 socio-economic data products can be dynamically searched, browsed, accessed and analysed via the AURIN portal. The significance of this approach is that urban geographers can access both the datasets and statistical analysis tools via the same online Portal and thus in theory, conduct their research more efficiently. In conducting our research we specifically focus on three practical use cases which utilise machine-to-machine access from the AURIN portal to the national census data maintained by the Australian Bureau of Statistics (ABS). The first use case involves location quotient analysis in New South Wales to analyse the financial sector. The second use case focuses on understanding economic growth across industry sectors across Queensland using shift-share analysis. The third use case aims to identify spatial autocorrelation between low income and other spatial variables in South Australia. Through this in-depth case study approach we have identified there are some strengths and weaknesses with the AURIN portal. We have demonstrated that the portal can successfully search, interrogate and visualise spatial-statistical data from across Australia. We have also demonstrated that the AURIN portal can successfully conduct simple spatial-statistical analysis all via a single online platform. However, we have also found there remains significant challenges in manipulating and visualising complex multi-dimensional datasets through the portal. Yet these challenges are not considered insurmountable and further research should endeavour to address them. In conclusion, it is important to highlight that online platforms such as the AURIN portal hold significant promise as effective planning and decision support systems which can be used to better shape our cities.

1. Introduction

We are now at point in time where more people live in cities rather than in rural areas and the global population is approaching 9.7 billion by 2050 (Department of Economic and Social Affairs, 2015). This is increasing the pressure on the infrastructure and services which support our cities. Advanced data-driven analytical and geographical visualisation platforms are needed so that multi-disciplinary teams can collectively work on shaping sustainable, productive and liveable cities and supporting rural hinterlands (Stimson, Tomko, & Sinnott, 2011). In an increasing digital world an online *one-stop-shop platform* which can support advanced spatial data analytics would be a valuable toolkit to support such endeavours (Tomko et al., 2012). The ability to turn data into information and ultimately knowledge (Ackoff, 1989) in a fast and effective way is an attractive proposition to urban researchers, policy makers and decision makers who are shaping the future of our cities. In this context, our research provides a critical analysis and evaluation of one particular online spatial-statistical platform by evaluating its ability to support three workflows of varying complexity, commonly employed by urban planners. However, the strengths and weaknesses identified are relevant to emerging online urban planning systems that comprise spatial-statistical analysis and visualisation services built on top of federated, open access, socio-economic data sets. Hence this paper's scientific value lies in its ability to inform the design of the next generation of online spatial planning and decision support systems.

In this paper we commence with a review the state of the art in online spatial statistical platforms. We then focus on the application and evaluation of the online advanced spatial-statistical platform known as the Australian Urban Research Infrastructure Network (AURIN)

E-mail addresses: c.pettit@unsw.edu.au (C.J. Pettit), Robert.Tanton@natsem.canberra.edu.au (R. Tanton), jane@itee.uq.edu.au (J. Hunter).
URL: http://cityfutures.net.au (C.J. Pettit).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.05.008
0198-9715/

portal (Pettit et al., 2015; Sinnott et al., 2015). We do this in the context of running the three user workflows using the Australian Bureau of Statistics (ABS) Census data based on (i) location quotient analysis (simple), (ii) shift-share analysis (moderate), (iii) spatial autocorrelation analysis (complex). Following the implementation of these three statistical analyses, we report on the findings, specifically the strengths and weaknesses of the AURIN online spatial statistical platform for analysing census data. We conclude by reflecting on the lessons learnt and the opportunities and challenges faced in building and using an online spatial statistical platform which provides machine to machine (programmatic) access to national Census datasets and advanced analytical and visualisation capability.

2. Online portals supporting spatial statistical analysis

A survey of methods for accessing census data to undertake spatial-statistical analysis, reveals the availability of a plethora of websites and portals. The United States Census Bureau provides its data openly via the Web (http://www.census.gov/). Data can be searched via their catalogue, downloaded as spreadsheets or programmatically accessed via one of their many Application Programming Interfaces (API). The Census Bureau also has a number of supporting data driven portals including, for example, their *OntheMap portal* (http://onthemap.ces.census.gov/) for exploring employment-related data. In the United Kingdom the Census data is also made freely available through the Office for National Statistics (ONS) and the data can be downloaded as a spreadsheet. Again, there is an open API that can be used to programmatically use and re-use the data. The United Nations also makes thirty-four of its databases available via a data portal (http://data.un.org/Default.aspx) and users can search and download the data as a spreadsheet or use their API to dynamically query the UN data. The Australian Bureau of Statistics (ABS) provide a similar census portal (http://www.abs.gov.au/). Data can be searched and downloaded in spreadsheet format and more sophisticated queries are supported by the ABS TableBuilder product which allows users to specify cross-tabulations of Census household or person level data. The ABS also provides its Census data through ABS.Stat (http://stat.abs.gov.au/), a Web service interface based on the Statistical Data and Metadata Exchange (SDMX) standard. However, all of these portals provide limited or no geospatial mapping and analytical functionality.

Advances in open and interoperable access technologies such as XML/GML/SDMX and Javascript, along with Web computing have seen the rise of Geospatial Cyberinfrastructures (Yang, Raskin, Goodchild, & Gahegan, 2010). Geospatial Cyberinfrastructures have come about as the next wave of online platform developments moving beyond the ability to merely search, discover, download and undertake simple mapping tasks. Geospatial Cyberinfrastructures offer advanced functionality whereby complex analytical routines can be undertaken using workflow engines supported by high performance computational and storage resources. However, the majority of effort in the development of Geospatial Cyberinfrastructures has focused on the domains of geology, environmental sciences, climate sciences, and coastal and ocean sciences (Pettit et al., 2013; Yang et al., 2010).

There has been a paucity of Geospatial Cyberinfrastructure applications built specifically for the domain of social sciences where spatial analysis and statistical methods are regularly applied. There are a number of data visualisation tools to support the creation of thematic maps from data associated with such census portals as the Thematic Mapping Engine (http://thematicmapping.org/engine/) and cloud based services such as the ESRI ArcGIS Server, but almost no applications which go beyond the creation of standard thematic mapping. The Statistical Online Computational Resource (SOCR) (http://www.socr.ucla.edu/) contains a library of standard statistical routines and data visualisation tools that can be run via a series of Web applets (Dinov & Christou, 2011). However, SOCR does not support spatial-statistical analysis and mapping and its technology does not support federated data search, discovery, analysis and visualisation. It is also important to note that there is a significant body of literature documenting the development and application of spatial planning and decision support systems (SDSS) with the like of tools like the desktop GeoDa tool which provides a powerful suite of spatial statistical, mapping and visualisation routines (Anselin, Syabri, & Kho, 2006). Specifically relating to spatial planning there is the series of edited book by Geertman and colleagues, which documents the state of play and adoption of such planning support systems (PSS) (Geertman, Ferreira, Goodspeed, & Stillwell, 2015, Geertman & Stillwell, 2009, Geertman, Toppen, & Stillwell, 2013). On review of this body of knowledge, it would appear that, to date, there exists only one online spatial-statistical platform that can support the end-to-end scientific inquiry process from problem formulation through to results publication. This online platform is known as the Australian Urban Research Infrastructure Network (AURIN) portal (Pettit et al., 2015; Sinnott et al., 2015).

Online portals such as AURIN, aim to provide social scientists with shared access to relevant federated datasets (via APIs) and spatial-statistical analysis and visualisation tools (via web services). They reduce duplication of data, software and computational resources and dispense with the need to purchase licenses to or download and install expensive data processing software on local desktop computers. However, there are a number of challenges associated with such shared, online e-social science platforms:

- Designing a single system and single user interface that can support the needs of researchers focusing on a very wide range of sub-disciplinary topics (e.g., demographics, housing, employment, socio-economics, transport, quality of life etc.) is a major challenge. The system needs to be simple and intuitive enough so that anyone can login and use it but also needs to support the advanced, complex data aggregation and processing requirements of sophisticated researchers. The three use cases describe in Section 5, evaluate and demonstrate AURIN's ability to support applications of varying complexity;
- Security is perceived as a significant issue by researchers using such online platforms. They need to know that: the datasets they are accessing; analyses they are performing; and maps/visualisations they are generating; are private, stored on secure servers and delivered through secure SSL connections. By leveraging the Australian Access Federation (AAF), AURIN supports single sign-on and user authentication, and ensures that users control access to their workflows and derived products, until they are ready to share or publish their results.

3. Methodology

In this research we focus on the development and application of a new federated data client to the AURIN portal which provides programmatic access to significant data resources to support online spatial-statistical analysis. This data client connects the rich datasets available via the ABS yearly Census directly into the platform and forms the basis from which urban researchers can conduct online spatial statistical analyses (Hunter et al., 2015). Given that the AURIN portal seems to be unique in its ability to support the urban researcher's scientific inquiry process end-to-end; it provides a suitable testbed for conducting our evaluation study into the performance of online environments for undertaking spatial-statistical analysis of census data.

In this paper we evaluate the AURIN portal's ability to support the urban researcher in conducting spatial-statistical analysis. The methodology is based on a use case approach whereby we illustrate and report on three use cases, one simple, one moderate and one complex spatial-statistical analysis. Each of these analyses is performed across three different Australian State jurisdictions, New South Wales, Queensland and South Australia. To provide a consistent method for reporting on and comparing each of the case studies, we have defined an urban researcher's typical 'scientific enquiry workflow' which is a

Please cite this article as: Pettit, C.J., et al., An online platform for conducting spatial-statistical analyses of national census data across Australia, *Computers, Environment and Urban Systems* (2016), http://dx.doi.org/10.1016/j.compenvurbsys.2016.05.008