

Analysis and Visualisation Tools for Spatially Integrated Social Science

N. Ward, T-K. Shyy, S. Irfanullah, F. D. Ungkara
University of Queensland, Brisbane, QLD, Australia

INTRODUCTION

The field of **Spatially Integrated Social Science (SISS)** recognises that much data that the social scientist examines has an associated geographic location (for example, a survey respondent's location). SISS systems use this geographic information as the basis for both integrating heterogeneous social science data sets and for visualising the results of analyses [1]. However, sourcing data sets, understanding relationships between the data and the geography, and implementing appropriate statistical analysis techniques are all time consuming and highly skilled processes. The UQ SISS System project aims to alleviate this burden from social scientists. The project is developing online tools that allow researchers to quickly access rich Australian socio-spatial datasets (e.g. voting outcomes and census data), conduct statistical modelling and visualize spatial relationships between the results. The software has three components that are described in more detail below:

- A repository of data and variables derived from Australian Bureau of Statistics Census Data and Australian Electoral Commission voting data;
- A set of statistical analysis and geospatial visualization services;
- A Web portal interface that enables researchers to analyse and visualize the data.

DATA AND METADATA RESPOITORY

The system incorporates data about Australian federal election voting results from the Australian Electoral Commission and Australian Census of Population and Housing data from the Australian Bureau of Statistics. It spatially relates variables derived from this data to

- Polling booth locations sourced from the Australian Electoral Commission;
- Statistical Local Areas, Local Government Areas, and Urban Centres and Localities region definitions sourced from the Australian Bureau of Statistics;
- Functional Economic Region definitions (sourced from the Centre of Full Employment and Equity (CofFEE) at the University of Newcastle; and
- Polling Booth Catchments region definitions sourced from the Queensland Centre for Population Research at the University of Queensland.

Figure 1 shows the workflow for ingesting these data sets and geographic definitions into the system.

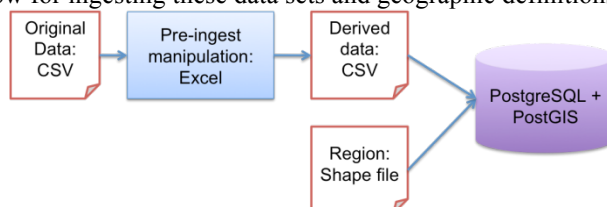


Figure 1: SISS Ingest workflow

The *pre-ingest manipulation* step in the workflow is currently an intensive, human-mediated process. Data analysts examine where to “join” the data sets and also create derived information relevant to the social science community. For example, during ingest our data experts add geographic region (such as polling booth catchment) identifiers to the derived data; and calculate voting *location quotients* by comparing the percentage of primary votes at each polling booth for a political party against the national benchmark percentage of the vote for that party. After processing, our system stores the derived data and region definitions in an object relational database (PostgreSQL) with spatial extensions (PostGIS).

Metadata about each of the data variables is also captured and stored in the database. Metadata includes the name of the variable, a description of what it is measuring, its source and lineage, its data type, and accuracy.

STATISTICAL AND GEOSPATIAL SERVICES

Our SISS system provides a service to generate thematic map displays of the voting and census data at varying levels of geography. The map generation service can classify the data using *equal interval*, *quantile* classification, or *natural breaks* approaches. A Java Servlet implements data classification and creates map layers served from an instance of Geoserver [2].

A statistical analysis service has been implemented using R and Rserve [3]. It supports regression, MANOVA clustering and principal component statistical analyses of the original and derived data.

The system exposes these tools as services that can be accessed independently of our user interface. This architecture also allows other (possibly separate) systems to call the services and provide their own visualisations of the results (see Future Work section below).

SISS WEB PORTAL

The SISS Web Portal provides a user interface for interacting with the mapping, classification, and statistical services described above. Figure 2a shows the result of a user zooming to a view of Australia, selecting to view variables associated with the Functional Economic Region level of geography, and asking for natural breaks classification of a particular variable. Figure 2b shows a map generated based on a statistical analysis of census and voting data that compares Brisbane polling booths that were frequented by single, generation Y, Greens Party voters at the 2007 Federal Election.

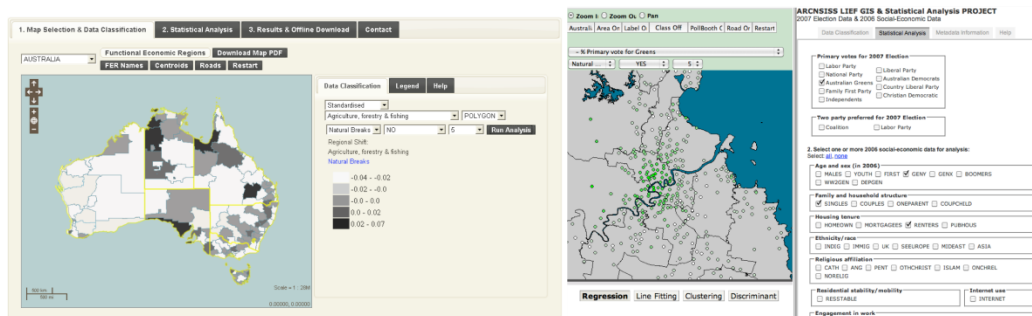


Figure 2: SISS Tools Screenshots: Classification and Statistical interfaces

Our web front-end interacts with the back-end services using web 2.0 technologies: jQuery and jQuery UI [4] access and display the classification and statistical services; an OpenLayers [5] JavaScript mapping interface allows interaction with the mapping service.

DATA SHARING, CONCLUSIONS AND FUTURE WORK

Although the original census and voting data is freely available, the “joined” data, the derived data, and some of the region definitions contain significant intellectual property developed by UQ and other institutions. For this reason, our data sharing approach has been to

1. not share the derived data and novel region definitions,
2. make metadata describing the derived datasets publicly available (through ANDS RDA [6]),
3. allow open analysis of the derived data through the web interface, and
4. allow negotiated access to the back-end mapping, classification and statistical services.

We feel this approach balances our obligations to protect the intellectual property of the research teams who created the derived data, while still promoting the data (through RDA), and allowing the socio-spatial research community to perform data analyses.

Although the current system only supports analysis of census and voting data, our services are generic, and can analyse and visualise other data sets. AURIN [7] have expressed interest in making our services available through their planned e-infrastructure. This would allow AURIN researchers to analyse the derived census and voting data, as well as apply our services to other urban research data such as economic and labour market data.

ACKNOWLEDGEMENTS

The SISS System project is a collaboration between the UQ e-Research Group [8] and the Queensland Centre for Population Research [9]. It builds on prototype SISS systems funded through from a mix of ARC Research Network and ARC LIEF funding. The SISS System project is one of five UQ Data Capture Projects funded by ANDS (the Australian National Data Service). ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative.

REFERENCES

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, 2 54: 63-88.
2. *Geoserver*. Available from: <http://geoserver.org/>, accessed 30 June 2011.
3. *The R Project for Statistical Computing*. Available from: <http://www.r-project.org/>, accessed 24 June 2011.
4. *jQuery*. Available from <http://jquery.com/>, accessed 30 June 2011.
5. *OpenLayers: Free Maps for the Web*. Available from: <http://openlayers.org/>, accessed 28 June 2011.
6. *ANDS Research Data Australia*. Available from: <http://services.ands.org.au/>, accessed 28 June 2011.
7. *AURIN (Australian Urban Research Infrastructure Network)*. Available from <http://www.aurin.unimelb.edu.au/>, accessed 30 June 2011.
8. *UQ e-Research Group*. Available from <http://itee.uq.edu.au/~eresearch/>, accessed 28 June 2011.
9. *Queensland Centre for Population Research*. Available from <http://www.gpem.uq.edu.au/qcpr>, accessed 30 June 2011.