

Exploring the Transparency of Government through Big Data on the Cloud

^{1,2}Prof Richard O. Sinnott, ²Yi (Noot) Fang, ²Edwin Salvador Pesantes

¹Melbourne eResearch Group,

²Department of Computing and Information Systems,
University of Melbourne, Australia,
rsinnott@unimelb.edu.au

ABSTRACT

There are many efforts exploring open data and the transparency of government across many countries. In Australia numerous initiatives are underway to expose government data that can allow for the transparency of government and the democratic process more generally to be better understood, e.g. [1-6]. It is still the case that much of this data is available through many independent web based sources and comparison across these is non-trivial. In this talk we will explore work that has been undertaken at the University of Melbourne through use the NeCTAR Research Cloud to integrate and analyse a wide range of web-based open data government resources. Two separate Master's level students undertook this work over a course of 10 weeks. The data sets that they used for exploring the openness of the Australian Federal government included: Political donations, Political Expenditure (2003 – 2013); Federal election results 2004 – 2013; Government contracts data 1999 – 2013; (Hansard) Political speeches 1996 – 2014; Lobbyists Details and Twitter. In total over 3,500 different data sets were integrated and explored on the NeCTAR Research Cloud utilising both noSQL technologies and big data analytics services (ElasticSearch, MapReduce).

Key scenarios that were explored were political tweeters: which politician tweets on which topics; who follows who (in/across parties); who is the most retweeted politician; the time when politicians tweet, and importantly the sentiment on particular topics tweeted by politicians and how these change over time, e.g. Carbon Tax, GFC, Education, Gonski etc. Other scenarios included political donations, e.g. which politician/party received the most donations and from whom; the correlation with political donations and grants and contracts that have been awarded; the analysis of parliamentary speeches.

This talk will include an exploration of these scenarios and cover the technical implementation details used for big data analytics on the NeCTAR Research Cloud. The talk will also cover further plans for extension and refinement of these solutions through on-going projects in the Melbourne eResearch Group.

REFERENCES

- [1] O. Definition. (2014). *The Open Definition*. Available: <http://opendefinition.org/>
- [2] O. G. Data. (2014). *Welcome to Open Government Data*. Available: <http://opengovernmentdata.org/>
- [3] Lindsay Tanner. (2010). Declaration of Open Government. Available: <http://www.finance.gov.au/blog/2010/07/16/declaration-open-government>
- [4] D. n. G. e. all. (2009). Engage: Getting on with Government 2.0 - Report of the Government 2.0 Taskforce
- [5] Open Data Government, www.data.gov.au
- [6] Open Data, City of Melbourne, <http://data.vic.melbourne.gov.au>

ABOUT THE AUTHORS

Yi (Noot) Fang is a Masters student at the University of Melbourne. His research dissertation was on the topic of a *Tweet Analysis of Australian Politicians*.

Edwin Salvador Pesantes is a Masters student at the University of Melbourne. His research dissertation was on the topic of *Promoting Transparency and Accountability of the Federal Government of Australia*.

Prof. Richard O. Sinnott is the Director of eResearch & Chair of Applied Computing Science at the University of Melbourne He was formerly technical director of the National e-Science Centre, UK; director of e-Science at the

University of Glasgow; Technical Director of the National Centre for e-Social Science and Deputy Director (Technical) of the Bioinformatics Research Centre at the University of Glasgow. He has published over 200 peer-reviewed papers in conferences/journals across a wide range of computing science areas with specific focus over the last ten years in supporting communities demanding finer-grained access control (security). He teaches Cluster and Cloud Computing at the University of Melbourne.