Submission for the 2015 IAOS Prize for Young Statisticians

Big Data and Semantic Technology: A Future for Data Integration, Exploration and Visualisation

Andrew Harwood and Andreas Mayer
Research Officers
01/10/1985, 29 and 02/07/1990, 24

Australian Bureau of Statistics
http://www.abs.gov.au/
Canberra, Australia
andrew.harwood@abs.gov.au and andreas.mayer@abs.gov.au

In a world of ever increasing data availability and user expectations, National Statistical Offices face mounting challenges to produce relevant and timely statistics. They need to transform their business practice to take advantage of big data – especially administrative data – by integrating non-traditional and survey data sources to maximise value, and utilising new technology to enable enhanced analysis. An example of a response to these challenges is the prototype LEED (Linked Employer-Employee Dataset) the Australian Bureau of Statistics is currently producing using semantic technology. A LEED integrates administrative tax data and survey sources allowing detailed microeconomic analysis. However, as data structures become more complex and multi-dimensional, data integration becomes difficult within traditional relational databases. Semantic technology allows for a flexible data structure, reusable classifications and standards, easy exploration of many dimensions, network analysis, and machine reasoning and inference on the dataset. The advantages of such an approach are demonstrated through two practical examples, showing how the prototype semantic LEED makes traditional data exploration and visualisation more effective, and how it enables new network analysis, to solve real business problems.