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

Andreas Mayer*
Australian Bureau of Statistics, Canberra, Australia – andreas.mayer@abs.gov.au

Andrew Harwood
Australian Bureau of Statistics, Canberra, Australia – andrew.harwood@abs.gov.au

Abstract

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 GLIDE (Graphically Linked Information Discovery Environment) the Australian Bureau of Statistics (ABS) is currently developing using semantic technology. This environment includes as a test case a prototype semantic linked employer-employee database (LEED) which integrates administrative tax data and ABS business register data 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 GLIDE makes traditional data exploration and visualisation more effective, and how it enables new network analysis, to solve real business problems.

Keywords: Semantic Web, Official Statistics, Data Visualisation