Census consultation – big data

The amount of data that is generally available is growing exponentially and the speed at which it is made available is faster than ever. The variety of data that is available for analysis has increased and is available in many formats including audio, video, from computer logs, purchase transactions, sensors, social networking sites as well as traditional modes. These changes have led to the big data phenomena – large, often unstructured datasets that are available potentially in real time.

Like many other National Statistics Institutes ONS recognises the importance of understanding the impact that big data may have on our statistical processes and outputs. So ONS established a Big Data Project to investigate the benefits and the challenges of using big data and associated technologies within official statistics.

Four practical pilot projects are included within the Big Data Project to give ONS first hand experience in handling big data. One of these pilots is focussed on gathering increased intelligence on addresses to improve the address register and make enumeration more efficient. The project has investigated the use of electricity smart meter trialselectricity data to model household unoccupancy patterns. Future research will focus on other data sources such as redirections data from Royal Mail.

Another pilot is considering the use of location data from mobile phones. It is speculated that such data might be used to either enhance or replace statistics about travel patterns for workers as typically produced in a Census. Alternatively such data could be used to produce statistics for different population bases such as the daytime population.

A further pilot involves the use of data from Twitter. Most tweets are public data and Twitter provides open source tools for accessing these data. Twitter provides an option for users to identify their current location. This means that tweets from a subset of users can be tied to specific locations over time. ONS has focussed on using these data to research the possibility of analysing aggregate mobility patterns such as migration around the country.

All pilots are at a research stage and ONS is looking at the potential advantages, as well as ethical and privacy concerns how to ensure that data protection principles and legal safeguards are upheld, these alternative data sources might provide over and above Census., as well as how to ensure that data protection principles and legal safeguards are upheld.