

**Professor Paul Longley, Professor of Geographic Information Science, University College London.**  
**“Passive collection – involuntary crowd-sourced data”**

#### **Abstract**

In an information age, more people are too busy to respond to social surveys leading to a decline in the reliability and accuracy of conventional survey techniques. Consumer behavioural data is being created as a by-product of everyday transactions. It is becoming a tradable commodity, however there is no open market place for trading of consumer data. The important ethical questions around the use of this data is being discussed at many different levels from individuals to academia and in the public and private sectors.

Spatial Data Infrastructures have an important role to play in anchoring the wide range of new information sources together.

Professor Longley finished by giving an overview of the research projects that are being worked on in this field.

#### **Notes to support the presentation:**

- CDRC seeks to populate some of the framework data, as such, this is a nice audience to talk to.
- Paul Longley as an academic viewpoint to give, which focuses on open software and solutions.
- Data is not just infrastructure (end user view), but should be seen as a resource. Data = the world's most valuable resource.
- Big data can come in many forms. This can offer both opportunities as well as threats.
- In the information age, many of us are too busy to respond to social surveys. This leads to very low response rates. Voluntary contributions to social surveys are in real decline.
- This leads to the threat that many conventional sources have become unreliable, creating a situation where we need Spatial Data Infrastructures to anchor new forms of data to a generated reality.
- The role of SDI is to anchor what we know (of those who leave digital traces of their identity) to what we know about them.
- Consumer data sources are a real share of digital data which is collected by our everyday transactions. They are a kind of exhaust which is not explicitly required.
- Data is a tradable commodity, volume of data can be seen as an instrument.
- There is no open market place for consuming consumer data.
- Big companies are thinking about building walls and silos, not recognising the value of the data collected.
- Change in ethics requires access to data from new sources which undermine the core aspects of data.
- CDRC have worked with many different companies from 2012-2016. These have mainly been on projects of mutual interest, which are proposed by academics. Some of these projects include:

- The roll out of smart meters has enabled domestic energy providers to have a continuous representation of data supply. This has made it easier to categorise energy users into four main profiles.
- Transport for London data on egress from underground stations, makes it easier to look at flows of information and people from footfall sensors. This enables a rethink in the different flows of people through workplace zones.
- Internet User Classification project, looked at identifying areas where people shopped online, and where people didn't. This led to better planning, and use of resources.
- Ethnicity estimator. Was a tool which allows cultural identity to be determined by analysing both forename and surname. The tool has over 2.9 billion names of people living. Each has cultural identifiers associated with it. Stated ethnicity is part of our own identification perception. E.g. John Barrow is unlikely to be Irish, whereas Patrick O'Mahney is. A model has to be used to predict this. The tool has led to a proposal for 56 different ethnicities to be included on the population census.
- Social mobility. Maps to show where people move from and their data characteristics.
- What is information?
  - Information is often costly to produce, but cheap to reproduce.
  - IT is easy to add value, through processing and/or merging information
- GI analysis is more about legal and ethical issues and the right appropriate uses of data by the private and public sector.

### Questions

- CDRC work close to privacy data, how do you get around it?
  - Ethics and accountability. There are usage advice from government. Understanding the rights of the research community to carry out research with this type of data.