Probes and Big Data: Opportunities and Challenges

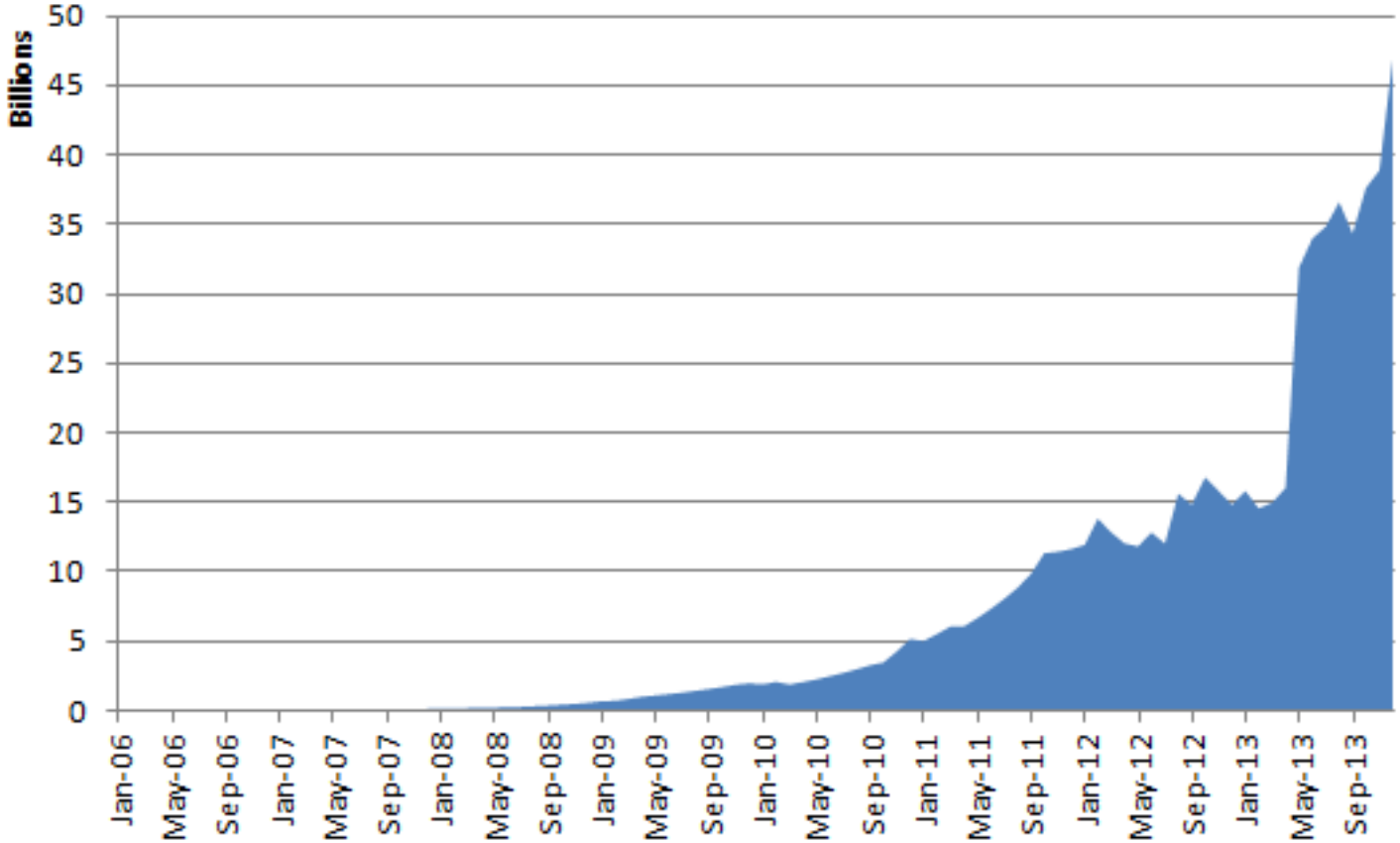
Fiona Calvert, Director Information Services and Mapping, Department of Transport Planning and Local Infrastructure

Department of
Transport, Planning and
Local Infrastructure

State Government Victoria

New data sources, probes and big data have the potential to revolutionise how we manage and plan the transport system

INRIX GPS data points – monthly, worldwide



Source: Danny Woolard, INRIX – Big Data in Transportation, BITRE Workshop Sydney May 2014

Probes and sensors are sources of big data, but not the only ones

Probes allow the movement of people and things to be tracked over the transport network

Key examples are GPS enabled devices and mobile phones (or cellular data

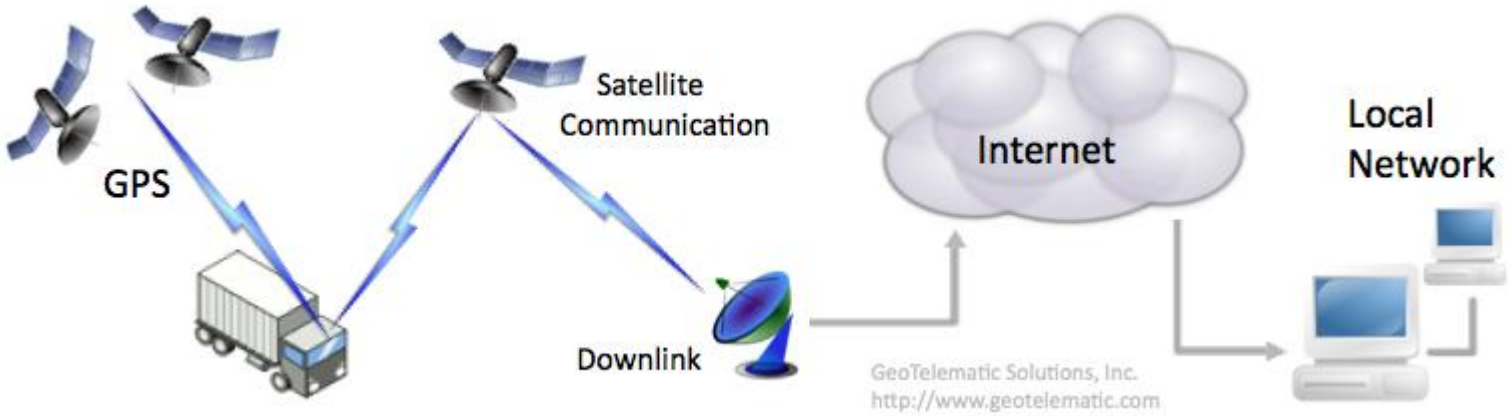


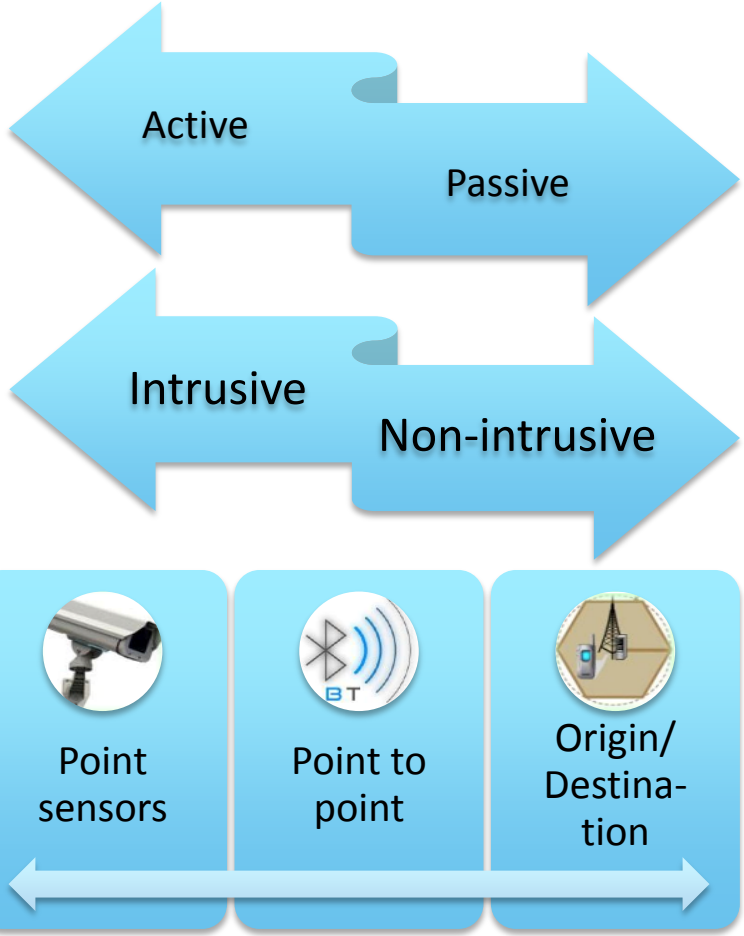
Table 1 Traffic data collection methods/technologies and information applications

Collection method	Data/information application: Active network management	Emergency response	Traveller information	Network planning
Conventional collections				
Road sensors				
Loop inductors	✓	×	✓	✓
Piezo-electric strips	✓	×	✓	✓
Pneumatic tubes	✓	×	✓	✓
Vehicle classifiers	✓	×	✓	✓
Household travel surveys (HTS)	×	×	×	✓
Traffic cameras	✓	✓	✓	✓
New technology collections				
Road-side sensors				
Infrared sensors	✓	?	✓	✓
Passive acoustic sensors	✓	?	✓	✓
Microwave sensors	✓	?	✓	✓
Lidar detection systems	✓	?	✓	✓
Video image detection	✓	✓	✓	✓
RFID[6]				✓
GPS-based HTS	×	×	×	✓
Probe-based collection technologies				
In-vehicle GPS	✓	?	✓	✓
GSM (cellular mobile data)	✓	?	✓	✓
Bluetooth devices	✓	?	✓	✓

Note Appendix B provides a more detailed breakdown of data collection technologies and traffic data collection items.

Source Bachmann et al. (2013) and BITRE.

There are many types of probes and sensors



Department of Transport, Planning and Local Infrastructure

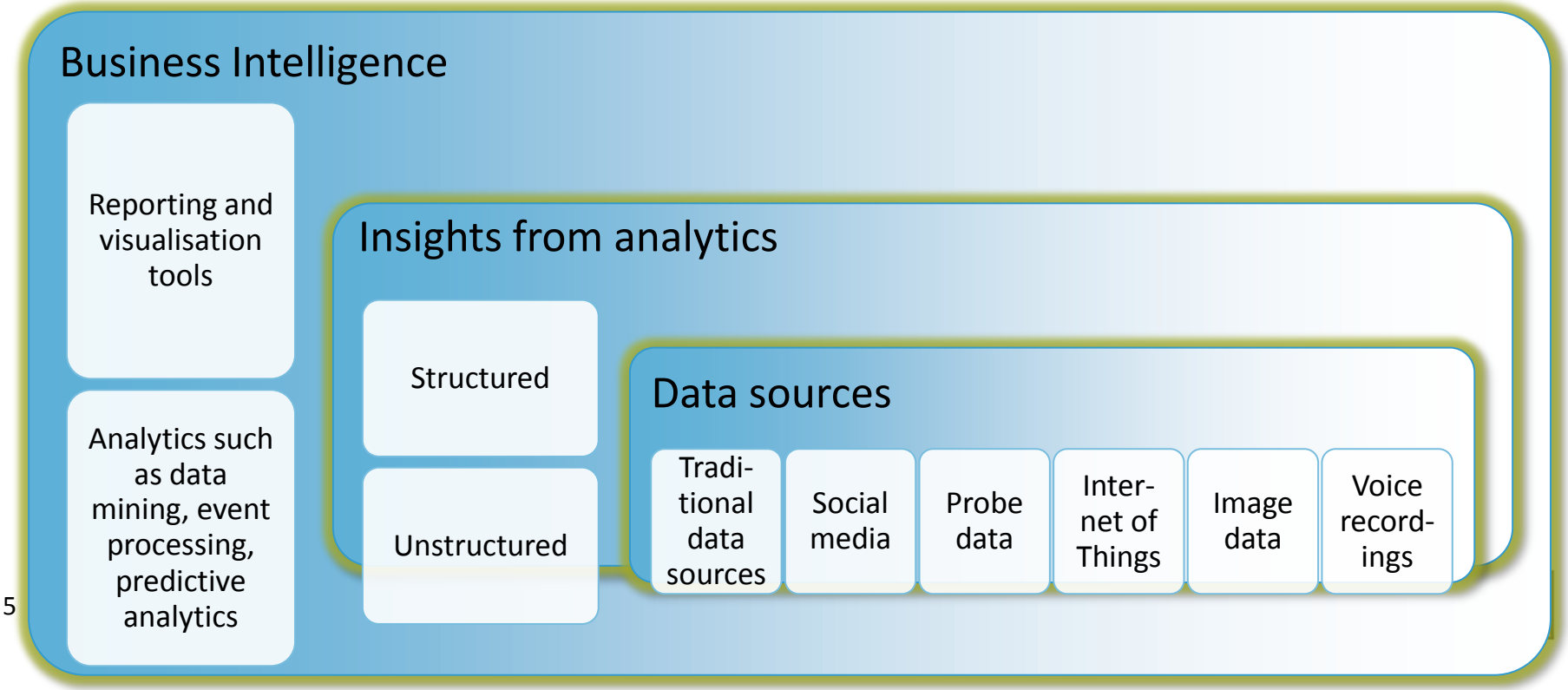
State Government Victoria

Big data is big because of its volume, variety and velocity

Traditional data is structured

New data sources include a wide range of unstructured data

'Golden nuggets' come from the insights provided by combining structured and unstructured data

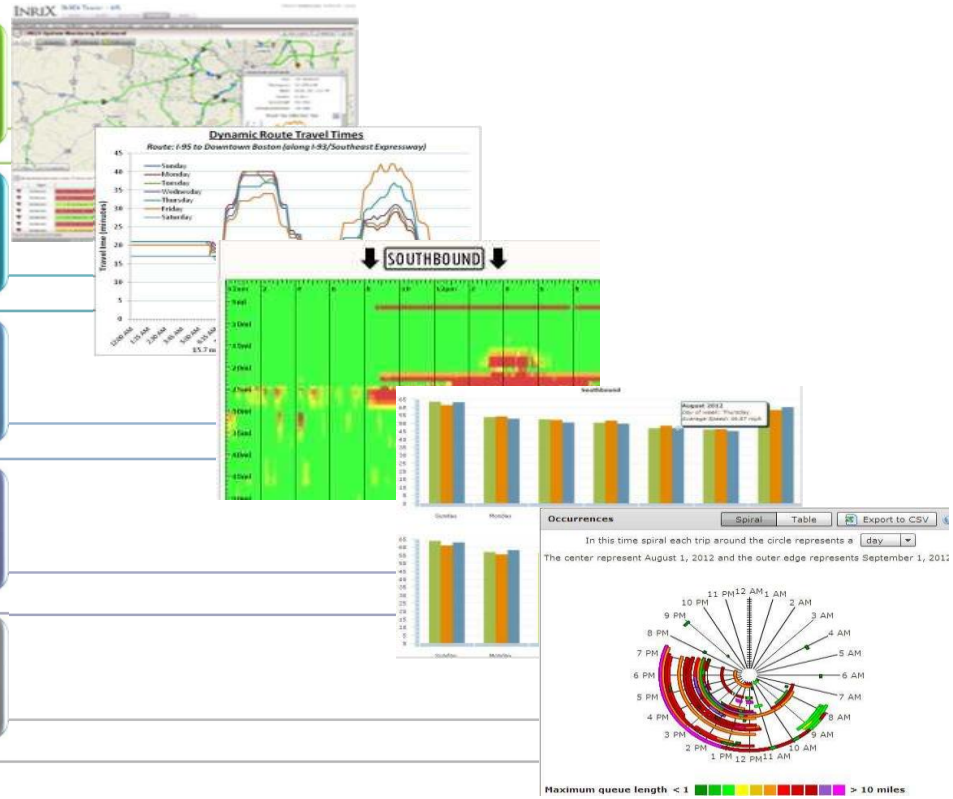


There are a range of uses of big data, particularly in measuring and understanding congestion

Data Analytics for Transportation

Turning Data into Information

- Real-Time Bottlenecks Identification
- Historical Traffic Download for Custom Analysis
- Congestion Tracking Analysis
- Congestion Trends
- Bottleneck Ranking & Movement



Source: Danny Woolard, INRIX – Big Data in Transportation, BITRE Workshop Sydney May 2014

Department of Transport, Planning and Local Infrastructure

State Government Victoria

Other uses include

Understanding demand

- Demand forecasting
- Marketing public transport
- Managing demand for road use

Operational performance

- Real time management including optimisation and incident response
- Network planning

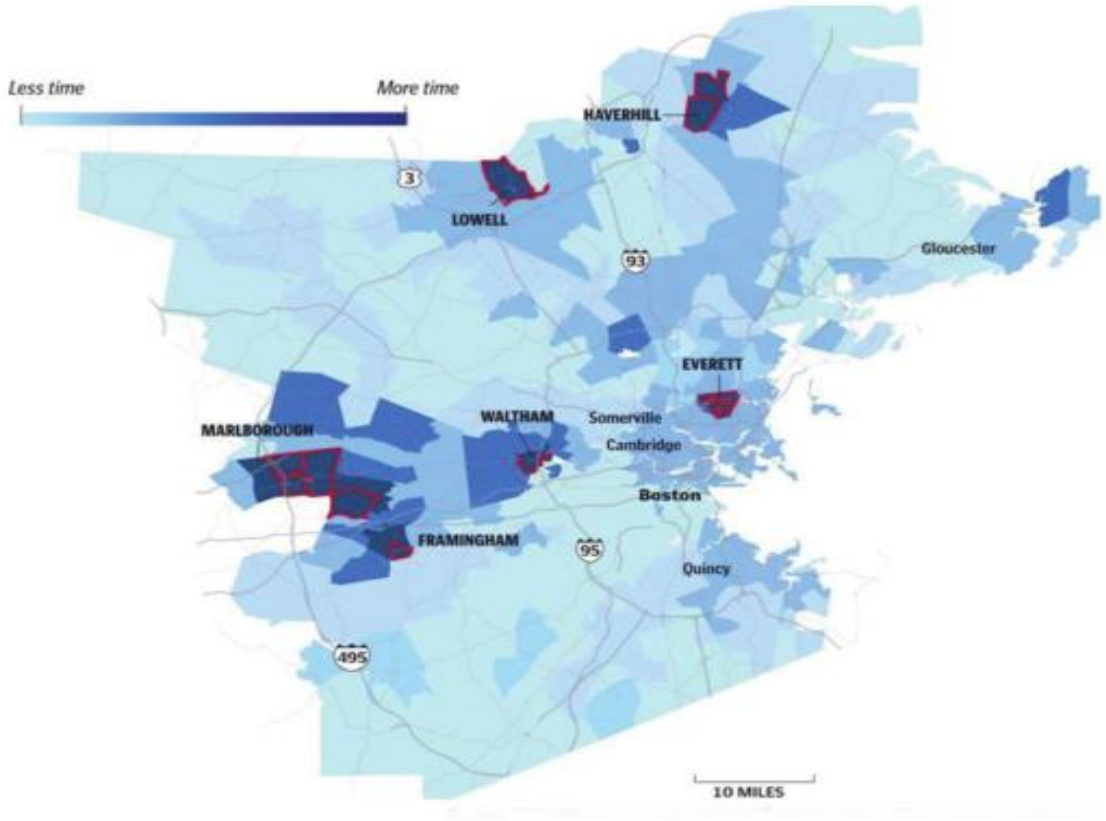
Customer service

- Sentiment mapping and crowd sourcing
- Customer service delivery planning and management (eg call centre resource planning)

Asset management

- Condition tracking (eg bridge and pavement cracking)
- Asset location and registers (eg speed signs)

Mobile phone data is beginning to be used overseas to understand transport activity



© Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., & González, M.C. (December 2012). Understanding Road Usage Patterns in Urban Areas. Scientific Reports.

Figure 6. Road usage based on cell phone data linked with census tracts to illustrate congestion exposure.

Source: FHWA Cell Phone Data and Travel Behaviour Research, Symposium Summary Report, Exploratory Advanced Research Program, February 2014

Department of Transport, Planning and Local Infrastructure

State Government Victoria

This example shows changes in taxi use in different traffic and weather conditions

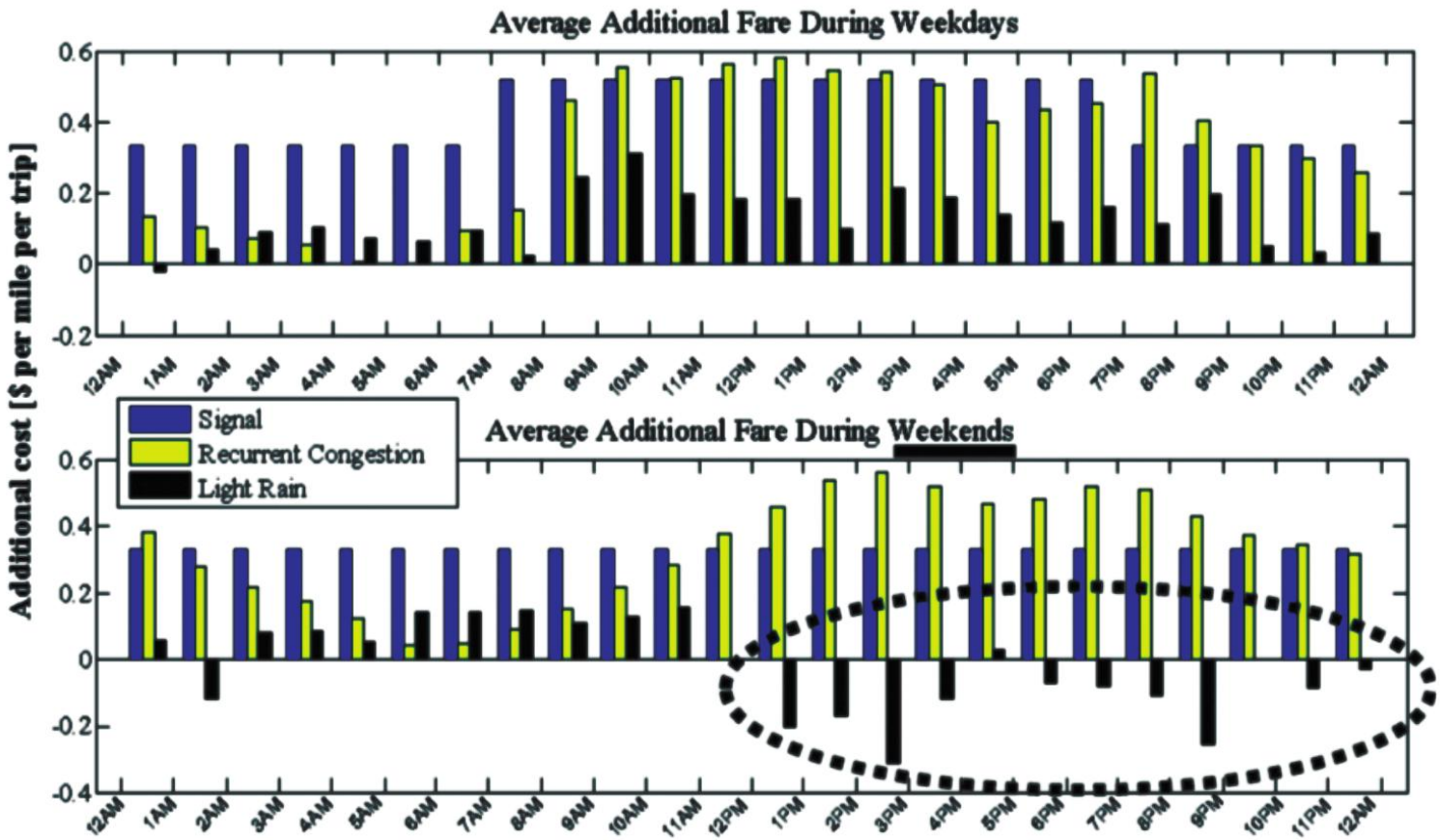


Figure 4. Taxi data used to determine demand for taxis during inclement weather.

Source: FHWA Cell Phone Data and Travel Behaviour Research, Symposium Summary Report, Exploratory Advanced Research Program, February 2014

Department of Transport, Planning and Local Infrastructure

State Government Victoria

Mobile phone data is also being used to understand origins and destinations of trips

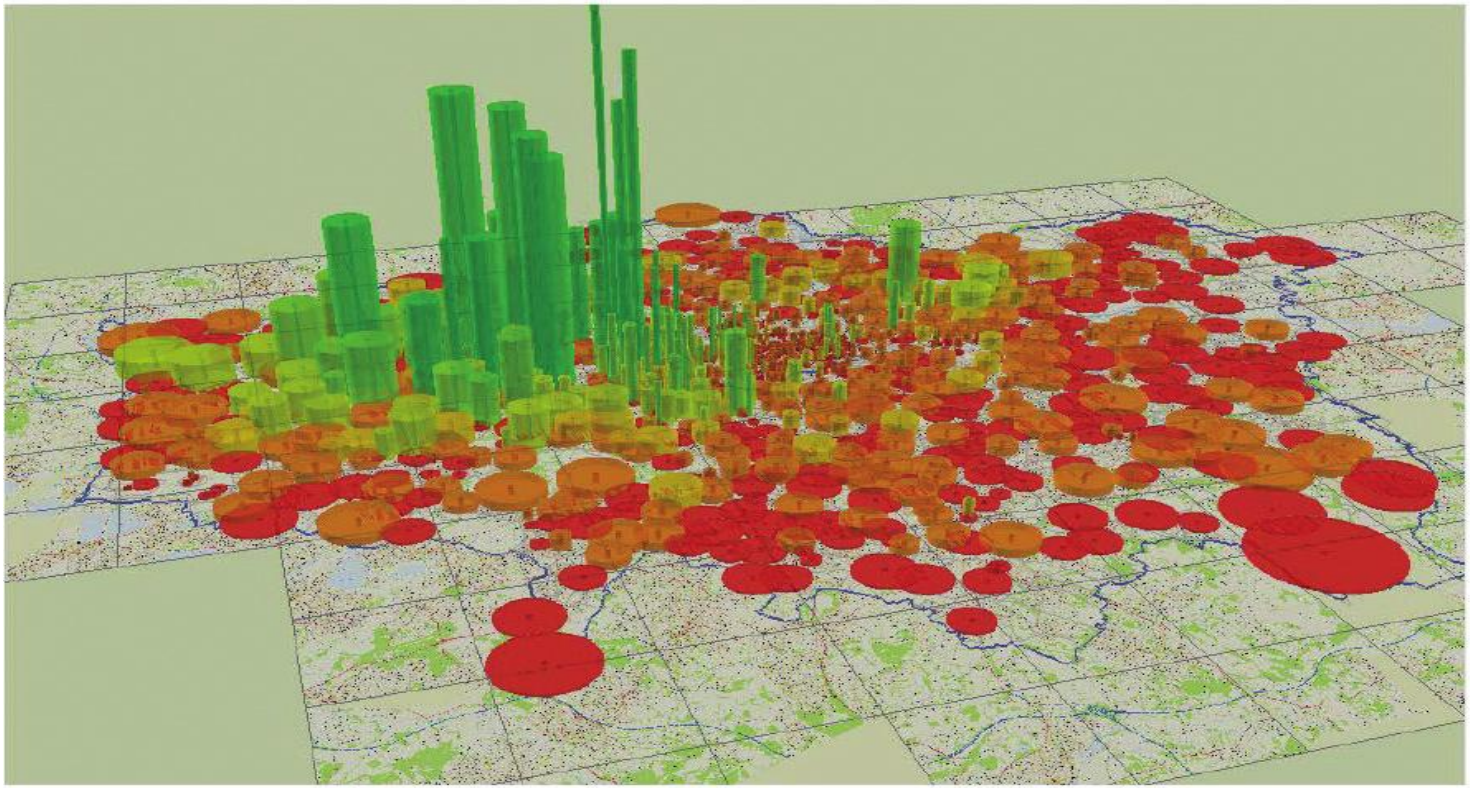
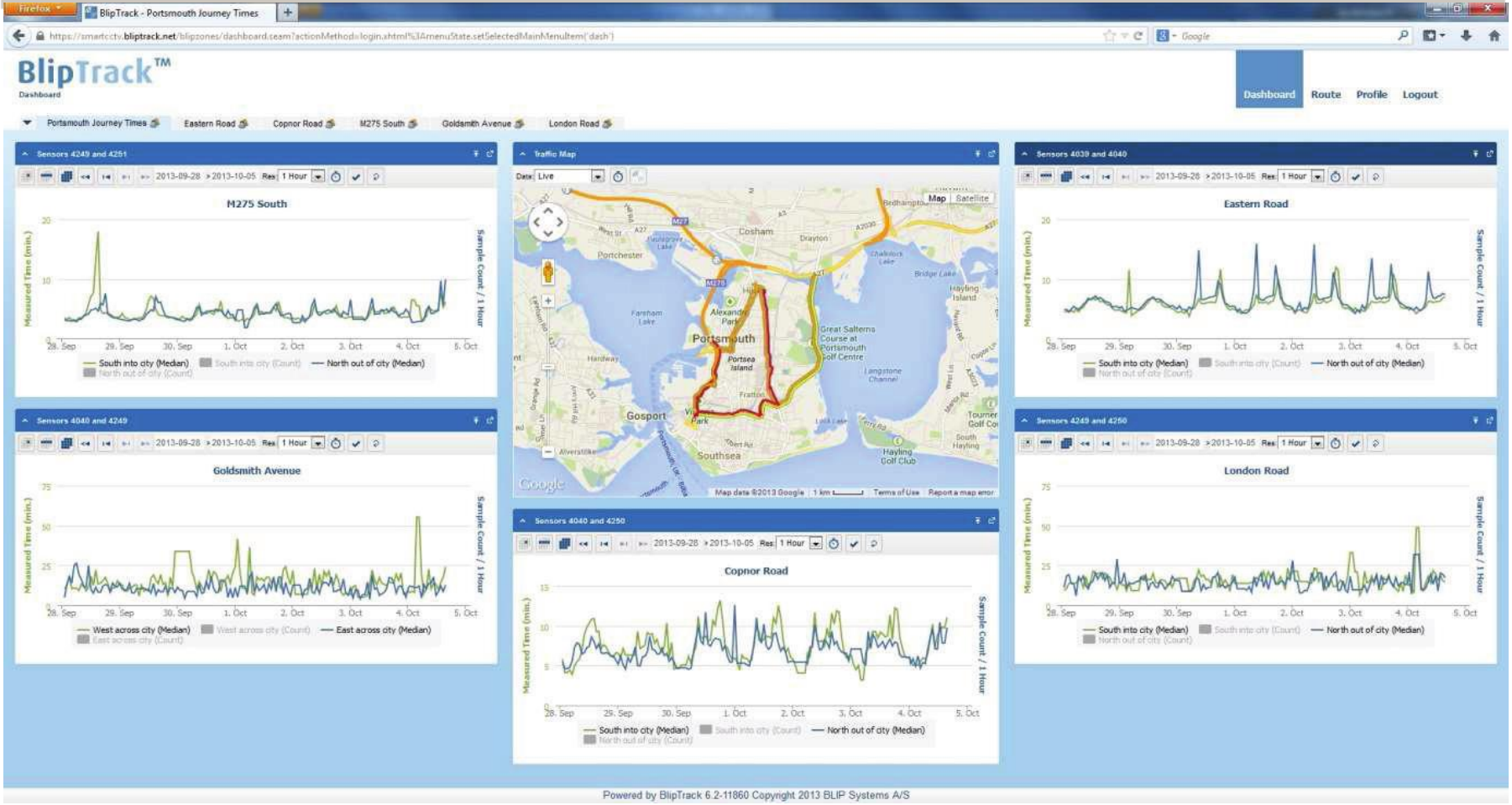


Figure 2. Cellular analytics illustrate where a region's visitors originated and highlights areas of demand.

Source: FHWA Cell Phone Data and Travel Behaviour Research, Symposium Summary Report, Exploratory Advanced Research Program, February 2014

Bluetooth and WiFi data is successfully being used to monitor and manage traffic, as in Portsmouth, UK

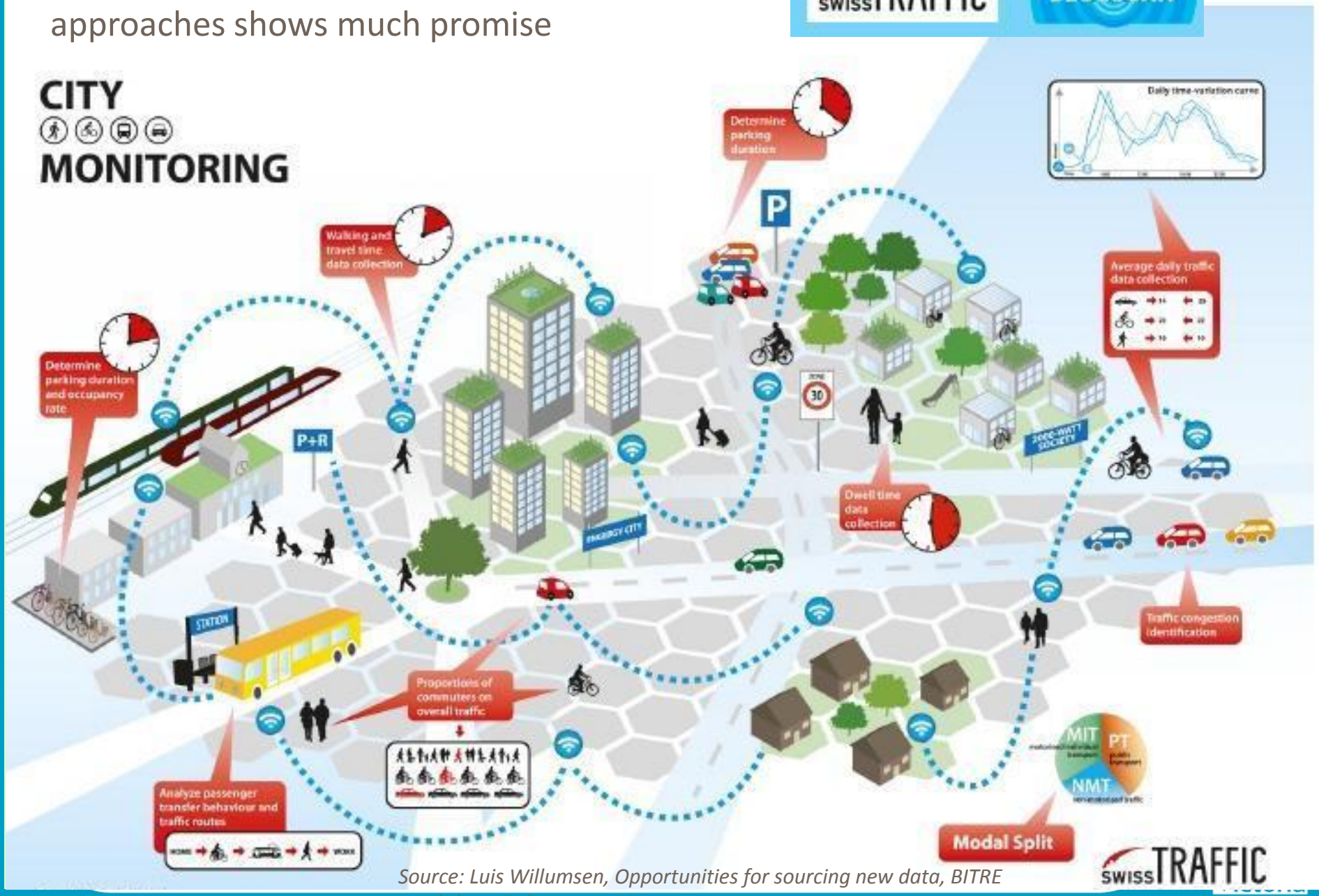


Source: SmartCCTV Ltd, Portsmouth Bluetooth and WiFi journey time and congestion monitoring system

Department of Transport, Planning and Local Infrastructure

State Government Victoria

A system approach combining multiple approaches shows much promise



Source: Luis Willumsen, Opportunities for sourcing new data, BITRE Workshop, Sydney May 2014

Part of the power of big data comes from its volume, which overcomes sampling issues

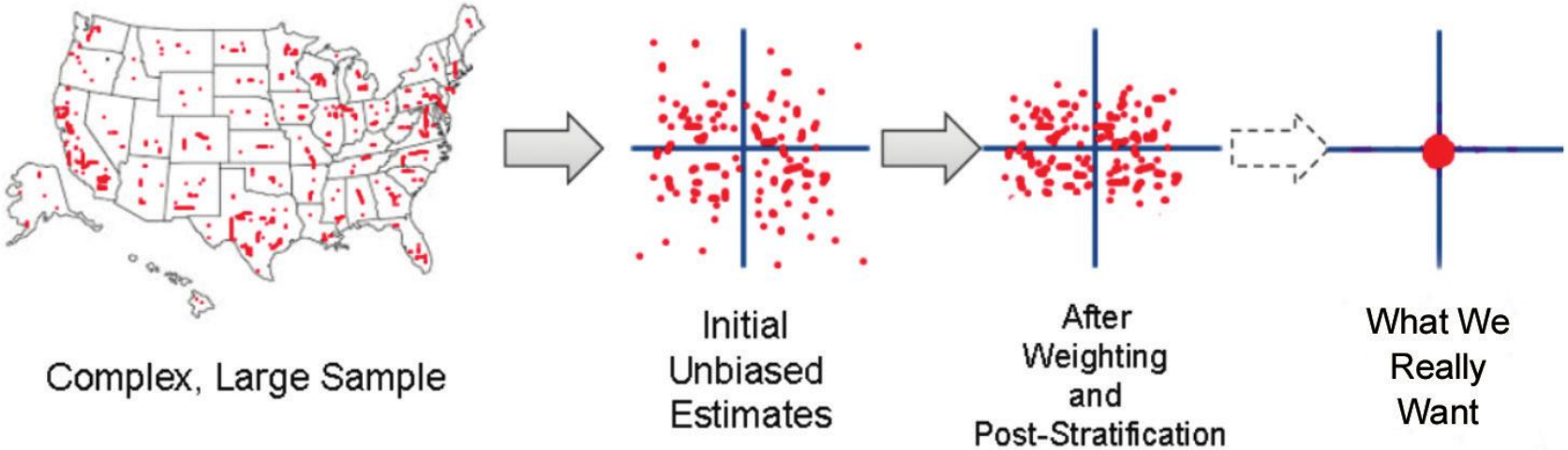


Figure 7. Example of a large national survey with sample sizes in the tens of thousands.

Source: Ben Pierce, Batelle, FHWA Cell Phone Data and Travel Behaviour Research, Symposium Summary Report, Exploratory Advanced Research Program, February 2014

Victoria is already involved

SCATS loops and traffic controllers

Bluetooth sensors in metro Melbourne (around 30 installed) and trialing analytical systems

Strava cycling and running data

Splunk trials on predictive maintenance for IT assets

SUNA data on traffic

Includes probe data from fleets using SUNA navigation systems

Image data analysis for speed sign locations, asset condition monitoring

Call centre quality control using voice recordings

Customer service centre resource planning

Sentiment mapping

Real time location information for buses and trams feeding passenger information displays and PT apps

800 000 hits per day for TramTracker

Key challenges are that big data changes the role of government and requires new capabilities and resources

Government goes from a data provider to a data customer

What is the role of government in providing a single platform and open data?

A fundamental change is needed in the way data is regarded

- Asset to be harvested

New capabilities

- Data scientists
- IT specialists and systems (eg storage, servers, software, skills)
  - eg Hadoop, Splunk etc

Questions we need to be thinking about include:

- What is needed in-house?
- What is needed for Government to be an educated purchaser of big data?
- How do we design our approach to get the most out of big data?
  - Avoid silos and duplication
  - Establish interoperability from the start

Privacy and security issues are a major challenge that must be addressed up front

Legislation differs state to state adding to the challenges

- NTC 2013 paper on Cooperative ICT recommends national principles/policy
- Victoria has just put in place a new *Privacy and Data Protection Act*, establishing a new Commissioner for Privacy and Data Protection

Issues include opt out clauses, user understanding of data use, potential for misuse….

Governments find it hard to take risks and care is needed so that in building an appetite for big data it is not seen as a sole answer to all information needs

We will need to take risks

- not every algorithm will work
- not every correlation will be meaningful
- not everything we think will be related will be found to have any link

Why Google Flu Is A Failure

It seemed like such a good idea at the time.

Source: Forbes, http://www.forbes.com/sites/stevensalzberg/2014/03/23/why-google-flu-is-a-failure/

Big data is not a panacea

- It is good at showing 'what' but not 'why'
- It will save costs sometimes but does not come for free

"Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

Source: Forbes, http://www.forbes.com/sites/stevensalzberg/2014/03/23/why-google-flu-is-a-failure/

Department of Transport, Planning and Local Infrastructure

State Government Victoria

The way forward involves pilots and proofs of concept, developing strong cases and partnerships

Partnerships

- InfraHack is a good example of the potential for partnerships
- Victoria has also started working with NICTA and the IBM Research Lab and is developing relationships with Google and HERE.com for GTFS feeds

Pilots

- Opportunities to build capabilities and establish proofs of concept
- Build business cases to change mindsets and secure resources

Partnerships across the transport portfolio can also be built to minimise costs and maximise opportunities

InfraHack 1st Prize: BetterWay

Product: 'road load balancing' tool to predict congestion and readjust traffic lights in real-time

Approach: combination of 'deep learning' (fast machine learning), predictive modelling and a graph-theoretic approach using datasets from VicRoads and TomTom

InfraHack 2nd Prize: LoZo

Product: 'loading zone finder' that alerts drivers to available loading zone bays as they approach the area of their destination

Approach: routing algorithms, PHP and Javascript using data from parking events from bays with sensors and Here.com API