

Big Data and Location – A Real or Imagined New Frontier?



Andy Coote reports on some of the insights gained from speaking and listening to some of the foremost experts in the field at the Strata 14 conference on Big Data in San Francisco and ponders the place of location in Big Data.

So, Big Data and Location, why should you care? In their recent report on Big Data, McKinsey (1) suggest it is becoming a key battlefield of competitive advantage, underpinning new waves of productivity growth, innovation, and consumer behaviour. One of the key application areas they highlight is geocentric personal navigation data. They assess the application of such data as being worth \$800bn worldwide during the current decade. Even if McKinsey are an order of magnitude too high in this forecast, it is still a staggeringly large potential market for the location industry.

So What is Different about Big Data?

Hasn't the location industry been effectively dealing with very large volumes of information for many decades?

The clue to the challenge is the abundance of data now available to commercial organisations and government. Mobile devices, earth observation satellites and the Internet of Things are just a few sources contributing to creating the world of Big Data. But it is about more than just Volume. Big Data also describes datasets with a high Velocity of change (such as real-time data streams), and with a wide variety of data types collectively known as the three V's (2).

This combination makes processing and analysis difficult using conventional tools. In particular, the volume and mix of structured and unstructured data, is a challenge for object-relational database management systems (such as Oracle and SQL*server) that most organisations currently use to underpin their data management. Here the major disruptive technology has been Hadoop, employed by search engines to produce the almost instant query response we have all come to expect from Google et al.

The huge additional business value to be derived from Big Data comes from what Accenture describe (3) as finding new insights. These might include identification of financial fraud, increasing retail sales or sources of inefficiency in government. None of these are new, but the science of what is often termed predictive analytics in Big Data circles, is introducing new tools and techniques which rely heavily on what we might have previously called spatial analysis and 4D visualisation.

Big Data comes with a torrent of new terminology which needs mastering to converse with the experts. The tag cloud below is a useful summary of some of the key terms, highlighting storage and analytics as the areas where location is critical. The boxed text provides some definitions.

Applications

According to John Morton, until recently with SAS but now an independent Big Data consultant, location figures in a wide range of applications because of its ability to reveal new information patterns and present information to senior executives visually. Some real examples were showcased at the recent Strata 14 conference on Big Data in San Francisco including:

Transport – Ian Huston, Data Scientist at Pivotal, sees Big Data analytics as a way to bring techniques from other disciplines, such as change point detection used in the wind turbine industry and cell population analysis from biology to complex problems of traffic management (4).

Retail – Susan Ethlinger, Altimeter Group, described an example of the use of location to identify problems in the supply chain of steak restaurants to illustrate deriving actionable intelligence from existing social and enterprise information sources (5).

Security – Ari Gescher, Palantir, presented “Adaptive Adversaries: Systems to stop fraud and cyber intruders”, where he described the use of geocoding of servers through IP addresses and various other “location assets” to provide intelligence to banks.

Health – genomics, the science of gene sequencing which involves very complex calculations on very large datasets takes centre stage in this sector. However, the medical insurers, such as Kaiser Permanente in the United States are also making heavy use of tools such as ArcGIS as part of their Big Data strategy.

Insights

Big Data's problem is that it's close to the top of the "hype cycle" of technology trends – so how do we discern the reality? A few gems from the gurus might help:

Geoffrey Moore, the author of *Crossing the Chasm*, suggests that 2015 will be the year for Big Data. Key for vendors is building the "minimum viable product" and establishing a beachhead in a substantial target market segment.

Ilya Sutskever of Google, suggests that little in data analytics has changed since the early days of artificial intelligence in the 1980s. There is much more data for training, faster computers, more storage and the name has changed (to machine learning). Fortunately the brain is still far more sophisticated than anything we can imagine building as a neural network.

Max Gasner of Salesforce.com, argues that there are no real general purpose predictive analytics platforms at the moment. He outlined the challenges for the Big Data industry and academics, as the need for BigML (a variant of XML for Big Data); a new user interface paradigm; and probabilistic programming languages. In short, Big Data lacks the solid scientific underpinning that Ted Codd provided for the relational database model, a generation ago. In other words Big Data is Waiting for Codd!

Emil Eifrem, CEO at Neo Technology, developer of Neo4j, which claims to be the most widely used graph database, and touted as the natural successor to the object-relational model, quoted predictions from Forrester research that 25% of organisations would be deploying such databases by 2017.

Location in Big Data Platforms

Different suppliers appear to have different views on the potential for location analytics in Big Data solutions.

SAP has taken the decision to embed Esri technology into the core of their product, which they believe will enable their users to leverage more simply geospatial tools as part of the HANA inmemory computing platform.

In contrast, Steve Jones, Cap Gemini, (partners with Pivotal in the Big Data space), believes the dominant approach will see designers building location analytics for their platforms as they find it useful. According to Jones, Big Data analytics will borrow the algorithms of GIS via good developers but will not try to "shoehorn" existing products into their architectures.

Another aspect of the Big Data debate was outlined by Steve Hagen of Oracle. Speaking recently at a UN GGIM meeting, he suggested that real-time feeds of location data are simply so huge that they are unmanageable in raw form and that filtering at source before loading into databases is the only viable solution. It seems to me however, that although deciding what to keep requires skills which geospatial practitioners uniquely possess, it does presuppose you know in advance what insights you might find.

Analysis

So much energy is being pumped into the Big Data story, it won't go away. Even if it is simply a rebranding of concepts that have existed for a long time such as business intelligence.

Why is it important to the location market? Because it is potentially a huge opportunity well over 50% of the presentations at the Strata conference used geocentric use cases to demonstrate their solutions or ideas. Furthermore, there seemed to be a general underestimation of the richness of insight that location analytics (what we used to call spatial analysis) could bring to the party.

Tailpiece

If you'd like to understand more about what Big Data means for the location industry, the AGI is organising an event on Tuesday 30th September in London titled simply "Big Data and Location". Hosted at the prestigious IBM Centre on the South Bank, it will bring together the main players from the Big Data and Geospatial worlds to explain technical concepts and showcase real applications. For more information go to the [AGI website](#).

References

1. [McKinsey Global Institute: Next Frontier for Innovation, Competition and Productivity](#)
2. [9 levers for Converting Big Data and Analytics into Results. Christy Maver, IBM](#)
3. [Realizing Data Value in the Insight Economy. Krista Schnell](#)
4. [Driving the future of Smart Cities](#)
5. [Social Data Intelligence: Integrating Social and Enterprise Data for Competitive Advantage. Susan Ethlinger](#)

This article was published in GIS Professional June 2014