



## “BIG DATA SOMETIMES NEEDS SMALL TOOLS”

Kevin MacDonald

kmac@loxcel.com

**“You can use the location components of your business data to gain valuable insight into your customers, your markets and new opportunities. But first you need to assess the quality and amorphous nature of the data, discard the chaff, and apply the right tools to what remains.”**

A few days ago, this email arrived in my inbox:

“Did you know that 70 per cent of all business data contains a location component? Put that information to work, and you can gain valuable insight into your customers, your markets and new opportunities.”

It was from Pitney Bowes, which recently acquired MapInfo and Group 1 Software, and was thus eager to do something in the GIS marketplace. At first, I thought the percentage was a bit optimistic. But then what is business data, but a record of all interactions a business has with its suppliers, distributors, customers – and all points in-between.

These interactions involve

the location of sales, donor addresses, franchisee locations, customer home addresses, shipping destinations etc. Any data that lacks a location component probably references other data that does.

If we accept that businesses are scared to purge data from their servers because they believe the future value from data mining accrues with the amount of data retained, and that storage costs continue to trend toward zero, the volume of business data with a location element should continue to grow unabated.

But how can a business (or any other type of organisation) extract value from this growing mountain of data that sits behind its firewall – without going on an expensive and time-consuming wild-goose-chase?

Wikipedia says that “Big data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times.” Exceptional technologies can be exceptionally expensive, so an organisation has to be exceptionally selective in the data it decides to pursue.

The mining industry faces similar prospects: a vast wilderness that obscures a rich but tiny mineral deposit. This industry uses aerial and magnetic surveys, and core samples to identify candidate sites, long before their “exceptional technologies” are employed to actually dig the mine.

Unlike rock formations, Big Data has a weak, amorphous physical form, which demands a more flexible approach to identifying candidate data for mining. Point-based data can be plotted on a map, which will reveal its location, topology and attributes. Domain experts can then explore what-if scenarios and apply gut rules, to either generate knowledge that can be applied directly to the business, or simply to assess the suitability of the data to further analysis by the “exceptional technologies” that IBM, Oracle, SAP and others have to offer. Other firms like eSpatial and Spatial Key offer lower-cost

visualisation solutions that model less-complex data in real-time.

When point data is well structured and pre-geocoded, even simpler, bespoke solutions become an option. An example is a project we recently completed at Loxcel. Our clients had asked us to produce a free-to-use map that could visualise in real-time all 420,000 radio transmitters across Canada.

Internet bandwidth isn’t infinite, so we chose a server-side solution to pre-cluster markers, in real time, in response to changes to the map’s filters, viewport position or zoom level. This map is hosted at <http://www.loxcel.com/radio> and currently uses Google as its base map.

But if Google licensing becomes prohibitive, we can deploy on other platforms like OpenStreetMap or MapQuest.

Pitney Bowes is correct: you can use the location components of your business data to gain valuable insight into your customers, your markets and new opportunities. But first you need to assess the quality and amorphous nature of the data, discard the chaff, and apply the right tools to what remains. 