

# Geospatial Big Data



Geospatial has always been considered as big data, both by its own advocates and many others, writes Ian Holt, a big data evangelist from the UK, in his latest column in *'GIM International'*. Yet large spatial databases and datasets are no longer enough to qualify as 'big data' as we now define it. Instead, they are just a piece of the broader picture.

Geospatial has always been considered as big data, both by its own advocates and many others. In fact, I could probably stop writing this column right now as we all collectively nod our heads. We could go one step further and find definitions and examples of big data, extremely large datasets and computational analytics. These terms will be familiar to anyone who has worked in GIS. It is not that long ago that big data was only a reflection of volume. For example, Ordnance Survey data was reported as being the largest (Oracle)

spatial database. As a representation of the entire visible landscape of Great Britain it certainly contains data on many features and is big by anyone's standards. However, whilst significant and vital for the UK economy, in its current form it is a masterpiece of data collection, storage and delivery. It is not 'big data' as we now define it, but rather – like other geospatial datasets – it is a piece of the picture.

If we are discussing geospatial big data we have to think more broadly and consider 'all' data. The important principle is not how you manage ever-increasing amounts of data in your silos; instead, it's more a matter of how you enable it to be linked to other data. With more and more satellite-based and sensed data becoming available, the data stores that we have built up are increasingly looking small and isolated. In fact, with the ever-increasing volumes of geospatial data, it has become apparent that current geospatial algorithms are not able to scale up so we need to research new ways of creating scalable ones.

Geospatial is merely one part of a much more ubiquitous data ecosystem. Let's take a look at the recent Manchester CityVerve project as a blueprint for smart cities. It needs not only geospatial data, but also many other types of data to provide everything from improved transport and healthcare to retail analytics and emergency-responder positioning. In many ways, big data can be seen as an umbrella term for a massive paradigm shift away from data collection/data delivery pipelines towards a knowledge-sharing environment – enabling the discovery of information through links and relationships. That's why knowledge bases such as Google's Knowledge Graph seem to be emerging in strength. These may be interpreted as evolving from linked data, but they provide a much more fluid knowledge repository without being rigidly built around a triple store.

Many have touched on the opportunities and challenges, but it is clear that geospatial big data is rapidly evolving towards data sharing as the industry responds by extending product ranges and capabilities. The traditional geospatial warehouse is being extended to support structured and unstructured data. Tools for geospatial big data analytics are emerging, such as visualisation, proactive location intelligence and data mining analysis. A key opportunity will be for the support of a geospatial big data service platform to complement the emerging 'Big Data as a Service'. This is not a winner-takes-all situation; this is how the geospatial industry can play the key role in underpinning big data.

**Ian Holt** has worked on a number of high-profile geospatial data and service implementations around the world, recently at Ordnance Survey and now as CTO at SplashMaps. He advocates the use of open technologies and standards to facilitate interoperable services. Ian volunteers for MapAction, providing services to help support humanitarian relief efforts.