THE ASSOCIATION FOR GEOGRAPHIC INFORMATION

Key Challenges and Opportunities Facing the Geospatial Information Industry: A view from the British Geological Survey

Kate Royse | British Geological Survey

'The move to monitoring geological processes in real time will mean that we will need to be able to ingest, process and visualise massive amounts of data. Significant changes in how the geoscience community handles geospatial data will have to change in the next five years'

The British Geological Survey has a bold new strategy to instrument the earth. What does that mean in terms of data and specifically geospatial information? In essence the survey is going to be moving from concentrating on the solid subsurface to monitoring processes below the ground in real time. This will necessitate not just the instrumentation of subsurface boreholes but also the use of satellite measurements and real-time monitoring of natural hazards.

The drive for this monitoring is to enable us to model, predict and mitigate the impacts of environmental change. To do this effectively will mean that we will need to look at novel ways to visualise and communicate our data to non-specialist users and combine our geological data with data from other sectors. Making use of mobile platforms and open data will provide us with many opportunities as well as some significant challenges over the next five years. These can be grouped into 3 areas discussed on the following pages.

Smart Data

Big data as a concept is defined around four aspects: volume, velocity, veracity and value. Although within BGS there are significant challenges going forward with volume and velocity in terms of data generation process and how to capture and store data. There are bigger challenges with the quality and usefulness of data. By 2020 big data will be the norm, the key to not being drowned in a big data deluge when the geoscience community opens the flood gates, will be the development of new technologies and architectures designed to extract value from large volumes of disparate data by enabling high velocity capture, discovery and or analysis. Currently the vast majority of data being generated is unstructured which means that more often than not we know little about the data unless it is somehow characterized or tagged. Metadata is not a new concept within the data information community but it is something that will need to become more prevalent if we are to make data 'smart' i.e. make data useful.

I believe that in the next five years we will need to move from a data management organisation culture to a learning organisation culture leveraging all the value behind the data that we collect and manage. Having lots of data isn't going to provide a solution to the 'big' science questions that we are faced with solving today. We will need to become better at filtering out the noise and holding onto the valuable data. Essentially the focus should not just be on collecting a vast amount of all the possible data out there but to contextualise each bit of data with its own specific context. Data needs to be understood and interpreted in a specific context for example what is the value of some information about a website visitor clicking on a link if the context that precedes and follows the clicking is not known? The capacity of big data to change how we do research will be down to our ability to make data smart.

Open Data

In the last five years we have seen an explosion in open data particularly in the public sector with the publication in 2012 of the government's white paper 'unleashing the potential'. The Open Government Data (OGD) initiative has emerged as a major movement in knowledge and data sharing. Although interoperability and accessibility are key challenges i.e. the governments linked data initiative, holding data in machine readable formats will no doubt become the norm in 5 years; there is a growing need to integrated social-economic and environmental data together.

Crowd-sourcing of data, through mobile devices, offers huge potential for fast acquisition of valuable datasets. The importance of crowd-sourced data in disaster resilience and response will continue to develop, given the rapid spread of mobile technologies and infrastructure across the globe. The role of new technologies and social media in the acquisition and sharing of data was highlighted by the 2010 Haiti earthquake. Communities affected by the disaster asked for help using social media. The result was that thousands of citizens were organised to collect, translate, and plot these pleas on maps so facilitating better-organised technical responses to the disaster situation. Data generated by apps (social media such as Facebook and twitter) and other background services leave huge trails of information that document every action a user makes, the full potential of this data is yet to be fully realised not only to improve and personalise web-services but also to improve communication and decision making at a national and local level.

'Cloud' technologies and services are challenging conventional models of data acquisition, management and processing. Despite widespread concerns over data security and caution over the role of private sector companies in managing public sector data, the rate of implementation of cloud services is growing rapidly. Several governments and state agencies are using cloud-based storage and processing services, e.g. the US National Institutes of Health uses Amazon Web Services to store its 200 terabyte human genome dataset from the '1,000 Genomes Project' (National Institutes of Health press release, March 29th 2012). The business model behind this initiative is one in which digital storage is provided free of charge but high performance computer processing is charged for. One consequence of this model is to make the data and the processing services accessible at low cost to all-comers radically reducing the barriers to entry.

Visualisation

Data visualisation unleashes geospatial data true impact and this is where we will see some of the biggest opportunities in the next five years, through improvements in data interoperability resulting in improvements in the visualisation of time series, point clouds and satellite data. The ability to do 4D visualisations in real time will allow us to harness the potential of geospatial data to become much more predictive, improving our forecasting and scenario planning. Data based policy making has necessitated the need for decision makers to be able to argue their case to a much more data orientated public. This necessitates improvements in not just how we communicate but also how we visualise data.

Visualisation of geospatial data has so far been flat; in two dimensions there is a significant push to start to integrate the subsurface with the surface. We are already seeing a growing interest to include subsurface data within BIM systems and integrating subsurface and surface infrastructure within smart cities. In the next five years the ability of technology to visualise seamlessly the flow of geospatial data and information from the surface to the subsurface will open up possibilities for managing the life cycle of buildings and infrastructure as well as monitoring and responding to the effects of environmental change and natural hazards within a virtual environment which will enable scientists to engage more effectively with policy makers, responders and the general public to provide effective and intelligent solutions to a wide range of problems and challenges.

Key Points

- How do we maximize the value of the data we collect and manage?
- How will new visualization technologies improve the value of big data?
- Crowd-sourcing of data, through mobile devices, offers huge potential for fast acquisition of valuable datasets but there are also limitations
- The 'Cloud' is changing how we acquire, manage and process data what new markets will this open up in 5 years time?

Conclusions

Within the Geo-Environmental sector there are significant challenges and opportunities around the development of smarter data, the use of Open data and cloud services and improvements in visualising the sub-surface and time-series data. There is no doubt that new approaches and technologies such as crowd-sourcing of data, mobile platforms and cloud services are radically changing the norms of data acquisition, management, processing and delivery. The biggest challenge is to be able to filter out the useless data and keep what's valuable. Data needs to be contextualised to be of use and there needs to be a shift in focus from collecting and managing to contextualising data and adding value. Finally the application of cloud based services and processing will significantly increase the efficacy of data sharing and processing which will encourage new and innovative ideas and technologies as the barrios to entry are lowered.