

Aggregating today's data for tomorrow's science: a geological use case (Invited).

Glaves, H.M., Kingdon, A., Nayembil, M.L., Baker, G., 2016.
British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK.

AGU Fall Meeting 2016. San Francisco, Ca, USA, 12-16 Dec, 2016. Presentation: IN44A-02. <https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/123405>

Abstract

Geoscience data is made up of diverse and complex smaller datasets that, when aggregated together, build towards what is recognised as 'big data'. The British Geological Survey (BGS), which acts as a repository for all subsurface data from the United Kingdom, has been collating these disparate small datasets that have been accumulated from the activities of a large number of geoscientists over many years. Recently this picture has been further complicated by the addition of new data sources such as near real-time sensor data, and industry or community data that is increasingly delivered via automatic donations.

Many of these datasets have been aggregated in relational databases to form larger ones that are used to address a variety of issues ranging from development of national infrastructure to disaster response. These complex domain-specific SQL databases deliver effective data management using normalised subject-based database designs in a secure environment. However, the isolated subject-oriented design of these systems inhibits efficient cross-domain querying of the datasets. Additionally, the tools provided often do not enable effective data discovery as they have problems resolving the complex underlying normalised structures.

Recent requirements to understand sub-surface geology in three dimensions have led BGS to develop new data systems. One such solution is PropBase which delivers a generic denormalised data structure within an RDBMS to store geological property data. Propbase facilitates rapid and standardised data discovery and access, incorporating 2D and 3D physical and chemical property data, including associated metadata. It also provides a dedicated web interface to deliver complex multiple data sets from a single database in standardised common output formats (e.g. CSV, GIS shape files) without the need for complex data conditioning. PropBase facilitates new scientific research, previously considered impractical, by enabling property data searches across multiple databases.

Using the Propbase exemplar this presentation will seek to illustrate how BGS has developed systems for aggregating 'small datasets' to create the 'big data' necessary for the data analytics, mining, processing and visualisation needed for future geoscientific research.