PropBase QueryLayer: a single portal to UK physical property databases

Andrew Kingdon [1], Martin Nayembil [1], Keith Holmes [2] and Graham Smith [1]

    (1)   British Geological Survey, Keyworth Nottingham NG12 5GG , UK and

    (2)   British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, UK

As the host institution of the National Geoscience Data Centre (NGDC), the British Geological Survey (BGS) holds significant volumes of subsurface data on behalf of the UK. These are derived both from the BGS's own data collection for research programmes, collected over many decades, and also as the national repository of data collected for other, often commercial, purposes deposited in the NGDC as a place of national deposit under statutory and other obligations. With the changing environment ensuring the security of the most basic requirements for human life, including the availability of potable groundwater, requires successful utilisation of finite resources from the subsurface, which in turn requires an ever greater understanding of the physical properties of the subsurface of the UK landmass.

BGS has already moved from a mapping to a modelling paradigm, where the "fundamental product" of all of BGS's outputs is no longer represented as a two dimensional (2D) paper map but as a computer visualised three-dimensional (3D) framework model. (Figure 1: an example 3D geological framework model). These show the geometric location of the subsurface information, realistically represented with shells or volumes representing geologically defined units. The next phase of this continuing process is to populate these geometric volumes with physical properties information that describe the heterogeneity of the subsurface.

Understanding the physical properties data is vital for the undertaking of the behaviour of the subsurface, and is directly relevant to understanding the composition and behaviour of the rocks and fluids underground. This is of enormous vast societal and economic importance, and of increase. Understanding of the subsurface heterogeneity of the UK, in particular the changes in attributes such as porosity and rock "strength" are of increasing importance in understanding the opportunities and threats represented by our subsurface. Once incorporated into the framework model this will be voxelated to demonstrate variation of property within the geometry (Figure 2).

BGS has for many years stored all digital scientific analysis and records in relational databases to ensure the long-term continuity of this information. However the structure of these databases is, by necessity, complex; each database, as well as containing positional reference data and model information, also contained metadata such as sample identification information and attributes that define the source and sample processing. Such metadata is critical to detailed assessment of the value of these analyses. It is however also hugely complicating for a simple understanding of variation of the physical property under assessment.

Given that the UK's populated areas are mostly underlain by clastic sedimentary rocks, understanding the variability of porosity is fundamental to understanding the nature of these rocks. However porosity data has been measured in a significant variety of ways, for a significant variety of end-uses over a long period of time. This means that the extraction of physical properties from these databases for a first look at understanding porosity is difficult; therefore the PropBase Query Layer has been created to allow a simplified aggregation of and extraction of all related data. The concept of the Query Layer is the presentation of complex data in simple, often denormalized, tables. The PropBase Query Layer brings together property information from various databases (each with its own database structure that reflects the nature of the data) into a single system. This means that data from all of the BGS's subsurface data holdings can be viewed together in simple interfaces.

Technical descriptions of the Query Layer (denormalized layer)

The PropBase data architecture is based around the concept of a query layer to present complex data in a simple but often denormalized set of tables and other programmable units within a relational database system. The query layer brings together property information from various databases each with their own relational structure into a generalised structure, so that there's a single consistent point of access of the data for any applications that may require the data. The query layer is implemented within an Oracle relational database system where the source databases also reside or are re-engineered into to facilitate easy loading of the data. The denormalization techniques used to build the query layer are not unique to Oracle and can be implemented on other RDBMS (Relational database management system). The query layer structure comprises

a set of tables, procedures, functions, triggers, views and possibly materialised views. The structure contains a main table PRB_DATA which contains all of the data with the following attribution:

- a unique identifier for each record
- the source of the data
- the corresponding unique identifier of the record from its parent database for traceability.
- the geographic co-ordinates of the record
- the depth values
- the type of property
- the value of the property
- the units of measure
- the appropriate qualifiers
- precision values and a full audit trail for the record

The data source, property type and units of measure are constrained by a series of dictionaries collated from the values used in the different databases from which data is extracted to populate the query layer. The property dictionary is a key component of the structure as this defines what properties and inherit hierarchies are to be coded and also guides the process as to what and how these are extracted from the structure.

The data model (Figure 3) shows that the structure also contains a child table PRB_DATA_COORD that holds secondary geographic co-ordinates in different projections from the primary record in the PRB_DATA table for a given record. This allows for the presentation of a property with its location in the primary projection and any others as maybe recorded in the database. The structure incorporates a level of flexibility because of the it's simplistic structure that enables us to add on any extra tables required linked off the main PRB_DATA table to capture extra attribution with a 1-to-many relationship or even a 1-to-1 if adding any extra attributes to the main table comprises the simplicity of the structure. In a similar vein to adding the capability to hold primary and secondary co-ordinates references for a record at different projections, the structure also incorporates a GROUND_REFERENCE table that allows secondary ground reference information at different surface level data to be recorded in a separate table to the primary record held in the main table. The surface level datum attribute in this extra table is constrained by a dictionary of such surface level datum types.

Given the size of the denormalized structure and the many property types and their values from various data sources, it's important that there's a co-ordinated technical approach to keep the layer synchronised. The query layer therefore makes use of oracle procedures written in PL/SQL containing the logic to carry out the data manipulation (inserts, updates, deletes) to keep the layer synchronised with the underlying databases. These procedures and/or packages are run as scheduled jobs at regular intervals (weekly, monthly etc.) or can be invoked on demand.

Implications for need to improve BGS database structures

Several databases have been in operation for 10-15 years without review. Work on PropBase has further identified redundancy within the data structures, data quality issues and opportunities for where improvement to the database structure that will not only allow delivery of information more effectively but also improve data quality at little cost.

Conclusions

The implementation of the PropBase QueryLayer has enabled BGS to find display and interpret more dataset with greater ease, massively simplifying the process of populating 3D framework models volumes with physical properties for parameterisation and study of geological intra-unit heterogeneity. This has enabled more rapid data discovery and population of 3D models with data held in our databases, enabling different datasets to be easily compared improving the data verification process. This technology will assist BGS is continuing to be one of the world leading national geological surveys.

Figure 1: Example 3D geological framework model of Manchester area (15 times vertical exaggeration) (Kessler et al , 2009)
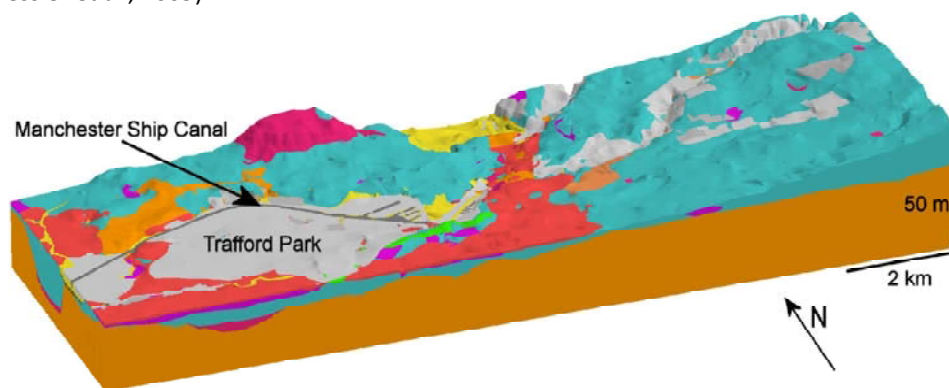

Figure 2 Example voxelated geological model parameterised with porosity data to demonstrate intra-unit heterogeneity.
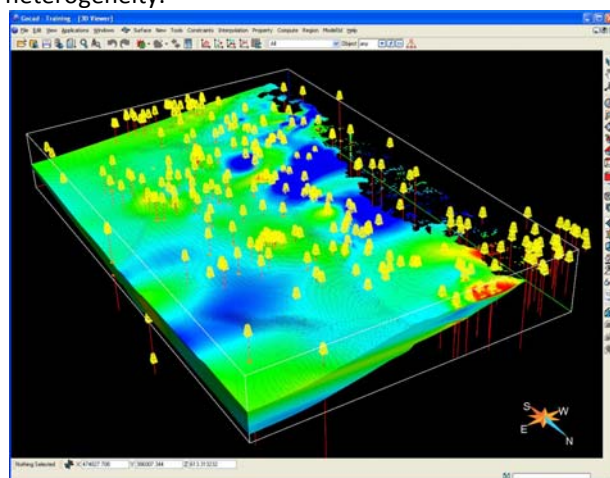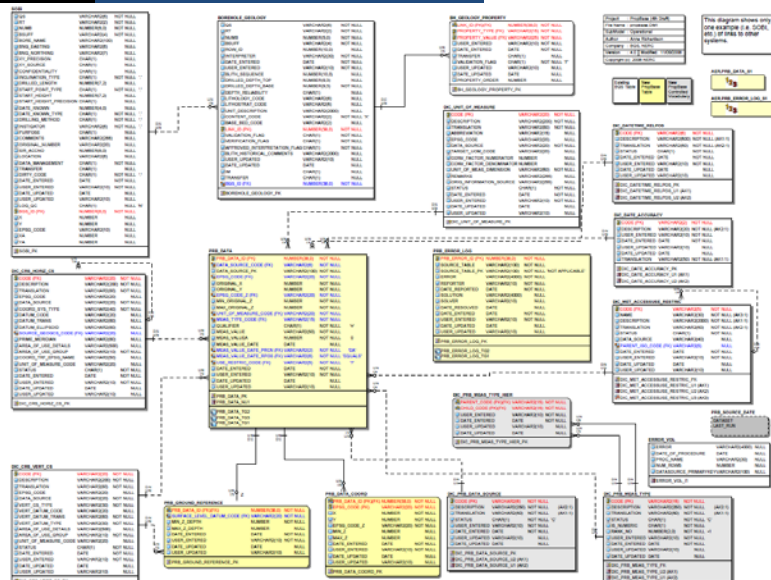

Figure 3: **Data model - PropBase Query layer**

References: **Kessler, Holger**; **Mathers, Steve; Sobisch, Hans-Georg.** 2009 The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. *Computers and Geosciences*, 35 (6). 1311-1321