Integrating commercially generated data into centralised geoscience data repositories for the benefit of urban environments

C. WATSON* and M. NAYEMBIL

Data Science Department – British Geological Survey, UK, cats@bgs.ac.uk
*presenting author

Abstract

The data model is often seen as a necessary step when producing information systems, they are rarely seen as a distinct output with their own inherent value and potential for future re-use. The geoscience community should improve how we communicate and share data models for past, present and future projects. This is of particular relevance when multiple stakeholders need to exchange and standardize information derived from a wide range of sources. This paper will consider the role the data model plays in the integration of commercially generated geotechnical data into centralized data repositories for the benefit of sustainable urban development.

1 Introduction

Geological Survey Organizations (GSOs) have traditionally focused on the creation and maintenance of geological maps that cover large geographical areas, many aspiring to map the entire country or region served by the GSO. In recent years many GSOs have moved towards a more strategic, targeted approach that involves focused studies of high priority areas, such as zones which are vulnerable to natural hazards and major cities. These targeted studies often involve the traditional geological map but also require a wider range of data sources to be considered. Relationships between geological properties and human processes need to be better understood which requires a greater understanding of interdisciplinary relationships, GSOs (and other public bodies) need to incorporate data from external, sometimes commercial, sources in order to see the whole picture.

1.1 Urban challenges

In recent years there have been a number of GSOs who have recognized that the complex and rapidly evolving nature of the ground beneath our cities presents specific challenges that require new approaches to data acquisition, data management and knowledge exchange. Multinational projects such as the COST action SubUrban (Campbell et al., 2014) have been established to assess the current state of play, document good practice and provide a suite of tools that should improve how geoscience data is communicated and used to support sustainable urban development.

Although city authorities and other stakeholders may recognize that the ground beneath their feet is a valuable resource and the density of data is often much greater than in lesser populated areas, it is often very difficult to collate all relevant information together in a useful manner. Another challenge is how geoscience information is communicated to stakeholders with such a wide spectrum of specialist knowledge. Information needs to be made available in ways which are appropriate to each type of consumer, from a geotechnical engineer carrying out a site investigation to a member of the public wanting to know if their house is at risk of flooding.

1.2 Data modelling

Arguably the biggest challenge GSOs face when attempting to understand urban subsurface environments is how to acquire, store and communicate all of the relevant data. In order to understand what a dataset represents and how it can be combined with other information sources requires that each dataset is well described and ideally well structured. Within computer science the analysis of data objects and their relationships to other objects is known as data modelling. Data modelling is often seen as task which needs to be carried out early in the development of databases and object-oriented programming. Data modelling should involve analysis of the data to be handled, learning about the real world objects the data represents and the relationships between these objects, this may be done very informally, however there are well defined formal techniques which are advocated by Steve Hoberman (2013) and others.

Despite advances in technology which have resulted in more data being made available in digital formats, there remains a large body of analogue data sources which are expensive to digitize. Financial constraints on GSOs and the increasing volumes and variability of data generated means that manual processes for subsurface data ingestion are unsustainable. In order to minimize manual processing it is necessary for newly acquired data to be captured and communicated between stakeholders using standardized digital formats that support automated processing. This can only be achieved if the data and the interrelationships between data are understood, therefore the data model is particularly valuable and by sharing them the community as a whole should benefit from greater levels of automation and standardization.

1.3 Geotechnical case study

The geotechnical engineering community provides one example of public and commercial organizations coming together to increase understanding of the ground conditions beneath our cities by developing common standards for data exchange. The community aims to increase data access, automate previously manual data acquisition processes and encourage data re-use. In the UK a format developed by the Association of Geotechnical and Geoenvironmental Specialists (AGS) is seen by industry as a way to reduce costs, lower transcription errors and speed up the building and engineering process by transferring data through electronic means (Bland et al. 2014, Walthall and Palmer 2006 Chadwick et al. 2006). The standard developed by the AGS committee is also referred to as AGS and unless otherwise stated the term AGS will be used to represent the standard. Pilot projects such as the Accessing Subsurface Knowledge (ASK) network operating in the greater Glasgow region, UK, are actively developing the community, workflows and digital tools needed to improve data sharing and increase knowledge of the ground conditions below the city (Campbell & Bonsor, 2013).

Building Information Modelling (BIM) is a term which has many definitions but most contain some reference to the need for the construction industry to improve communication amongst all stakeholders and advocate the use of new policies, processes and technologies.

The UK government announced in 2011 their intention to require "collaborative 3D BIM (with all project and asset information, documentation and data being electronic) on its projects by 2016 (The Government Construction Strategy, published by the Cabinet office). The BIM methodology is not confined to the UK and is becoming increasingly important globally. To-date BIM has focused on the structures which are being built and on the subsurface on which they need to be built, the geoscience community need to raise the visibility of geoscience data for use in BIM solutions.

Unfortunately the grand visions of open, free sharing of data between stakeholders, promoted by the creators of standards such as BIM and AGS, are rarely fully realized. One reason is that people interpret standards in different ways, in particular software vendors that create tools to support standards re-interpret the standard and produce implementations which are not completely compatible with tools developed by other vendors. This is particularly relevant to any organization, such as a GSO or city authority that needs to collate data from a wide range of data generators and standardize them in a single communal data store.

2 Integrating data sets

In order to deliver the technical aspects of the ASK Network vision required the development technical and cultural solutions to ensure the most relevant data for the community was shared in an efficient manner for the benefit of all. The responsibility for developing a technical solution fell largely on the BGS whilst Glasgow City Council enforced the cultural changes necessary to ensure contractors provide data in a standardized format. The **G**lasgow **SPE**cification for data **C**apture (GPEC) was developed as part of the project and was used as the standard data exchange format, GSPEC is essentially AGS version 3.1 with some additional rules (Campbell & Bonsor, 2013).

A data flow was developed which allowed authenticated data donors to submit GSPEC files through an online service, this triggers an automated validation process and, if valid, files are transferred to the communal data store. This data workflow was initially designed in 2013 and a solution was launched early 2014 that closely resembles the draft design shown in Figure 1.

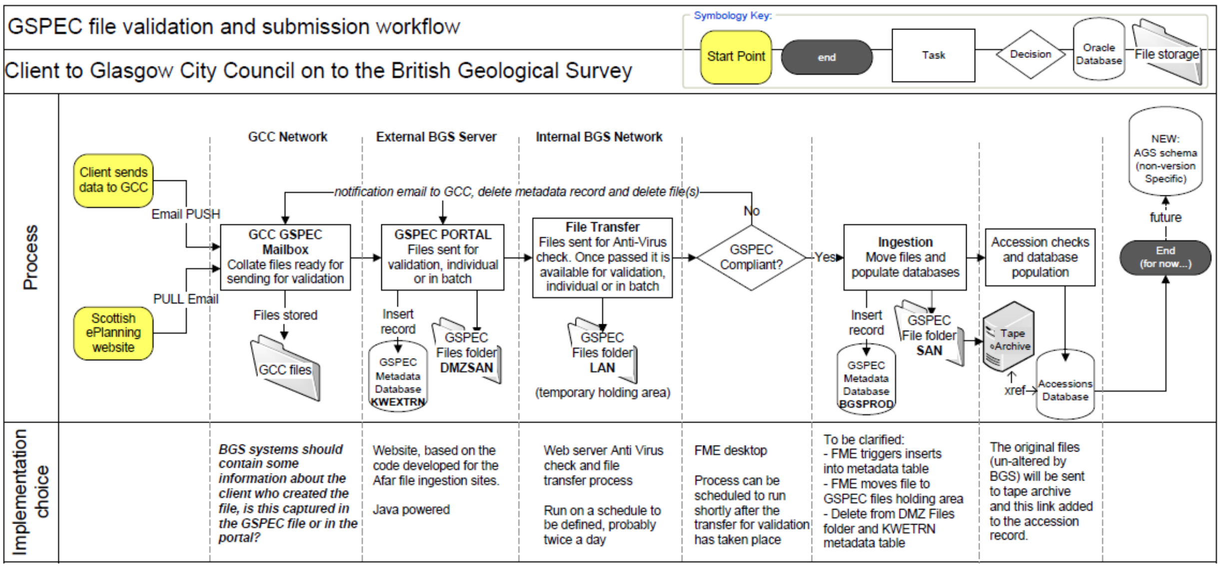

Figure 1: Draft workflow for ASK Network data acquisition

2.1 The role of data models

Data models were crucial throughout the ASK Network project and beyond, they were used to inform data donors what was expected of them. Through existing AGS documentation, AGS software and additional knowledge exchange activities by BGS and GCC, contractors were given clear definitions of what data formats they were contractually obliged to use when supplying data to GCC. A data model was created as part of the design process when building a relational database to store the GSPEC/AGS data which was subsequently used to populate the BGS national geotechnical database and form the basis of 3D geological models of Glasgow.

In related work, the visualization software GeoVisionary (Napier, 2011) has been extensively used to view borehole data from a wide variety of sources, many with custom data formats. It was decided that a new data model (figure 2) was required that could describe all of the common borehole information required by GSOs and commercial organizations, source data could then be transformed into this common 'omniborehole' format for visualization in GeoVisionary. This work was carried out in a relatively short period because the data models most of the source data formats were publically available.

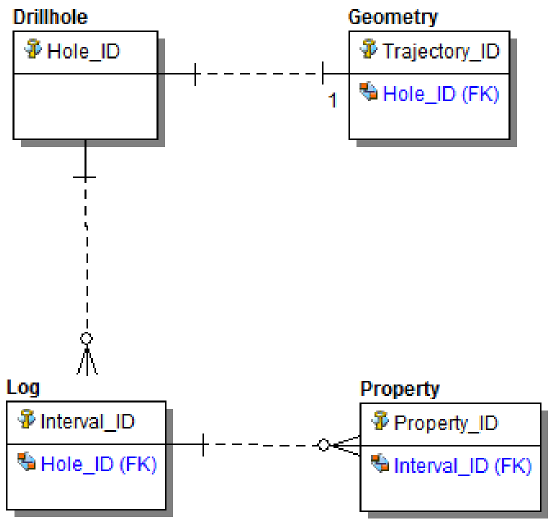

Figure 2: Top level data model diagram for boreholes

The BGS have also recently begun a project with Keynetix, the software company behind the most popular geotechnical software in the UK, HoleBase. This project, called BIM for the Subsurface, aims to provide the common data exchange formats, validation tools and systems architecture necessary to integrate the geotechnical data from GSOs into commercial geotechnical & 3D BIM software (Grice & Kessler, 2015). HoleBase users will be able to browse and download BGS held geotechnical data, 3D geological models, or parts thereof, for use within ground investigations and related BIM projects. At the end of an investigation users will be able to submit new geotechnical and interpreted model data back to the communal data store. During the early phases of this collaborative project the open sharing of data models between the project partners has helped to inform the design of existing and future databases as well as the exchange formats and web services required to pass data between commercial and public organisations.
Both the GeoVisionary and Keynetix work have highlighted how GSOs are subject matter experts that can use data models to communicate data requirements with software specialists, who may be experts in their own field but have little understanding of geological concepts.

2.2 Results

The GSPEC data model was intentionally open and accessible, this was achieved by informal communication with public and commercial organisations supported by formal documentation and automated validation checkers. This has had a tangible impact on the way geotechnical data is shared in Glasgow and the wider area raising the profile of AGS across Scotland and the UK.
The AGS data acquired through the ASK network is combined with existing lithostratigraphical borehole data, geological maps, digital terrain models and other observation data to produce a suite of 3D geological models. The ASK network community can download a copy of the 3D model for analysis on a local PC or interrogate the model via an online tool which allows users to produce synthetic boreholes, cross sections and horizontal slices across the modelled area, Figure 3.
At an ASK Network workshop in March 2014 Alex Scott, a project manager for Scottish Water commented "Using the ASK Network 3D models from BGS has increased certainty at an earlier stage in project planning, improved estimating and financial forecasting, and given us greater confidence in option selection in infrastructure upgrades in Glasgow.". Whilst Jackie Bland, Geotechnics Ltd and Chair of the AGS Data Management Working Group commented "GSPEC is AGS format really being used in the way it was originally intended to be."

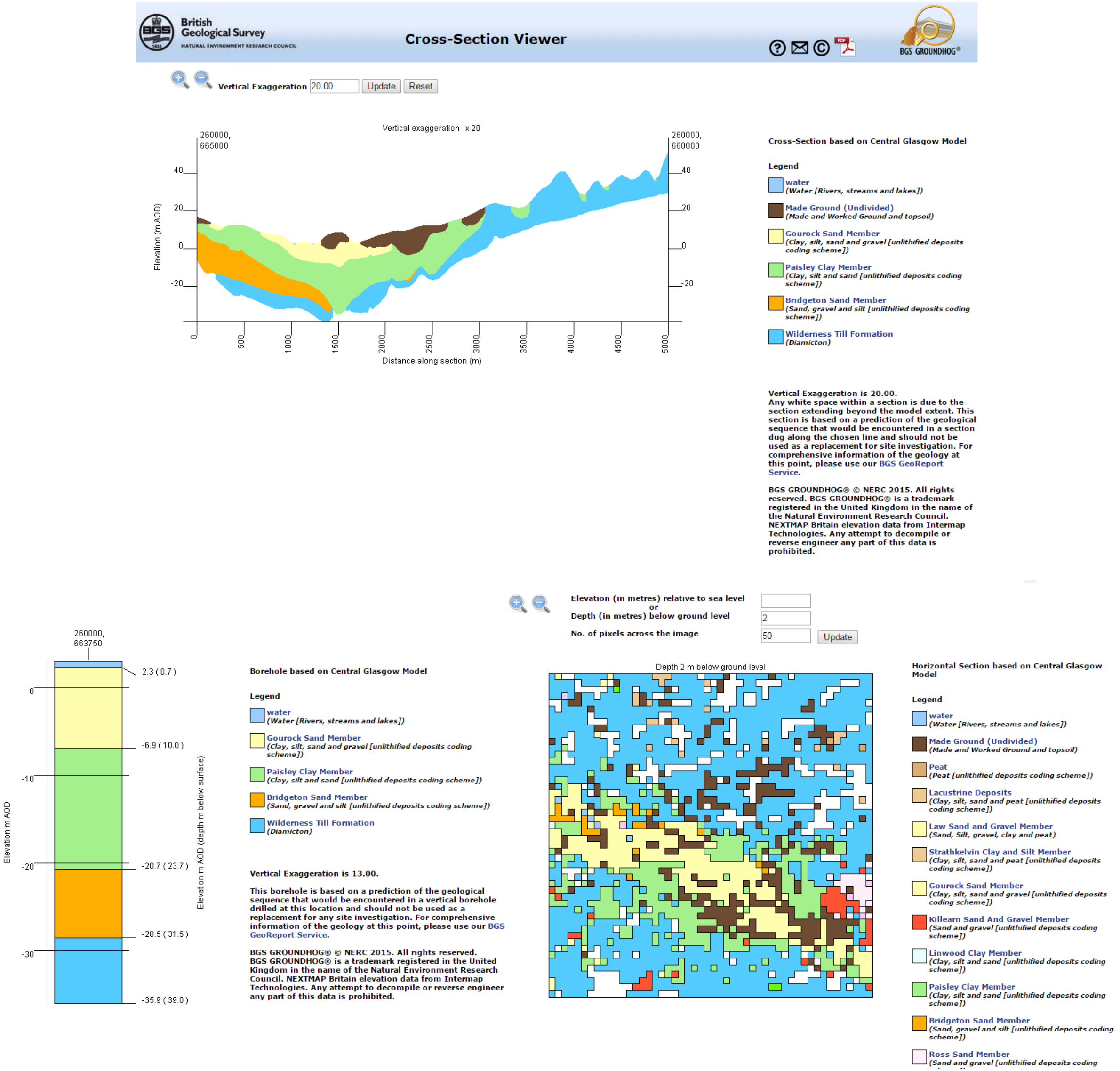

Figure 3. Selected screen shots from the synthetic borehole, cross section and horizontal slice tool for the Glasgow 3D geological model

3 Discussion

There are no hard and fast rules covering how data model sharing should be carried out yet there would appear to be an appetite for the creation of a public geoscience data model sharing resource. One attempt to fill this gap was the data model library available at the website www.EarthDataModels.org (Figure 4) which was created as part of the NERC funded knowledge exchange project Open Geoscience Data Models (Watson et al., 2014). The earthDataModels website contains a series of public, free to use and modify data models collated from a range of GSOs and commercial companies, covering topics such as boreholes, geotechnical data and metadata. Data models can be documented and shared in formal and informal formats, the earthDataModels website promotes the sharing of data models in all formats with the house style being a semi-formal package of documents, consisting of high level entity-relationship diagrams and plain English descriptions of the data model. The descriptive documents focus on the core attributes necessary to represent the data for each topic covered and include tips on how alterations could be made to the design to meet local needs. Where possible implementations of the data models are published along with sample data so that users can see how the data model works in a relational database such as Oracle or Microsoft Access.

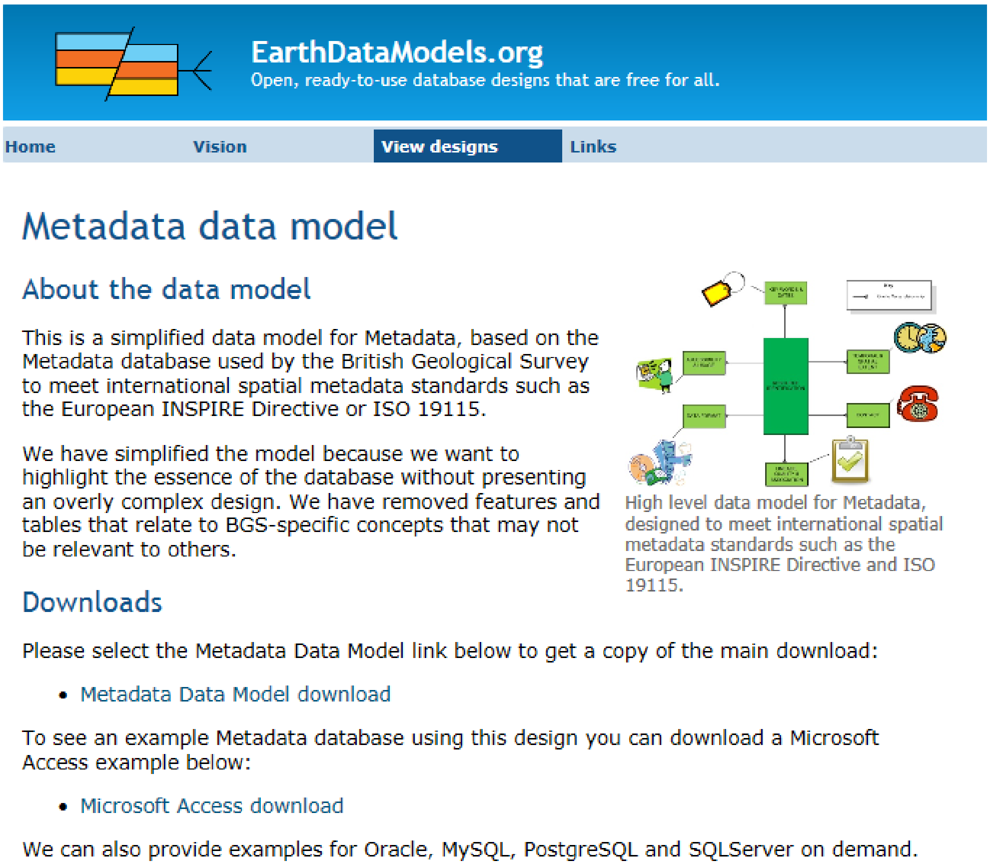

Figure 4. EarthDataModels.org metadata data model

4 Conclusions

Data models are extremely helpful as mechanisms for communicating the meaning of datasets and therefore support the linking of previously unconnected sources of related information. Data models being seen as a necessary step in designing most information systems, they should also be considered as an output in their own right.

Sharing our understanding of the structure and meaning of datasets through data models will advance our ability to standardise and integrate previously unconnected information sources, this will directly support the ambitions of GSOs and city authorities who aim to improve understanding of the ground beneath our cities.

Data model sharing can occur in a wide variety of ways, from visual diagrams of conceptual data models to technical specifications of physical data models in XML schemas, each method has its advantages and disadvantages. It is more important to recognise the value of data models and encourage their re-use than worry about the exact method of documentation and communication.

References

Bland, J, Walthall and D. Toll (2014). The development and governance of AGS Format for geotechnical Data. Proceedings of the 2nd International Conference (ICITG). Information Technology In Geo-engineering. IOS Press BV, Netherlands. 67-74.

Campbell D., J. de Beer, D. Lawrence, M. van der Meulen, S. Mielby, D. Hay, R. Scanlon, I. van Campenhout, R. Taugs, I. Eriksson, M. Bacic (2014). Sub-Urban - Transforming relationships between geoscientists and urban decision-makers (COST Action TU1206). Proceedings of the 14th World Conference of Associated Research Centers for the Urban Underground Space (ACUUS 2014), Seoul, Republic of Korea, 24-26 September, 2014, 284-289.

Campbell, D. and H. Bonsor (2013). The ASK Network: Glasgow Specification of Data Capture. http://nora.nerc.ac.uk/506350/

Grice, C. and H. Kessler (2015). Collaborative Geotechnical BIM technologies. Presentation. http://nora.nerc.ac.uk/510823/

Hoberman, S. (2013). Data modeling made simple with ER/Studio Data Architect. Technics Publications. ISBN: 9781935504481

Napier, B. (2011). GeoVisionary: virtual fieldwork for real geologists. V1 Magazine. http://nora.nerc.ac.uk/13558/

Walthall, S and M. Palmer (2006). The development, implementation and future of the AGS data formats for the transfer of Geotechnical and Geoenvironmental data by electronic means. GeoCongress 2006: pp 1-4 doi: 10.1061/40803(187)109

Watson, C., G. Baker, and M. Nayembil (2014) Open Geoscience Data Models: end of project report. British Geological Survey Open Report (OR/14/061) (Unpublished) http://nora.nerc.ac.uk/508791/