British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Open Geoscience Data Models: End of project report

Informatics Programme

Open Report OR/14/061

BRITISH GEOLOGICAL SURVEY

INFORMATICS PROGRAMME

OPEN REPORT OR/14/061

Open Geoscience Data Models: End of project report

Carl Watson, Garry Baker and Martin Nayembil

Keywords

Report; Data Models, Knowledge Exchange, Database Design.

Front cover

Project Logo.

Keyworth, Nottingham British Geological Survey 2014

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk

Tel 0115 936 3143 Fax 0115 936 3276
email enquiries@bgs.ac.uk

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3241 Fax 0115 936 3488
email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

Tel 0131 667 1000 Fax 0131 668 2683
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090 Fax 020 7584 8270
Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE

Tel 029 2052 1962 Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800 Fax 01491 692345

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF

Tel 028 9038 8462 Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

Website www.bgs.ac.uk
Shop online at www.geologyshop.com

1. Executive Summary

This report describes a three year knowledge exchange project, OpenGeoscience Data Models, which was funded by the Natural Environment Research Council (NERC) Knowledge Exchange programme. The project was aimed to encourage an open sharing of geoscience data models amongst a community of Geological Survey Organisations (GSOs), industry and academics. The data model is a key part of successful information management because it provides a centralised description of the meaning and inter-relationships of the information. An online repository for data model designs would be created, providing access to open, ready-to-use database designs that are free for all. The project would attempt to build a self-sustaining community that would develop these designs, and contribute additional designs, to meet the needs of the wider environmental community.

The stated objectives of the OpenGeoscience Data Models project were:

- Support NERCs Environmental Monitoring objective of 'Enabling environmental data sets to be applied in new commercial and societal services and products'
- Contribute to the ongoing global improvement in the quality of geoscientific data
- Contribute to the ongoing efforts to ensure water security
- Improve efficiencies within the global community of geoscientific data owners and consumers
- Generate economic opportunities for the BGS and UK plc

The majority of time spent on the project was directly related to the technical aspects of the work, adapting existing data model designs for a wider audience, producing introductory and technical documentation and creating the websites to disseminate these outputs. Knowledge exchange communications with project partners took place regularly and has continued after the project officially ended. Special technical and community building visits helped to cement relationships between community members whilst presentations at international conferences raised awareness of the project aims and outputs.

By the end of the three years the project team had delivered on each of the stated objectives, had produced two separate web resources, one branded as part of the popular BGS website and another designed as an independent community website. A large amount of plain English documentation was produced and data model designs were all made available for free download via the two project websites.

New contacts and knowledge exchange relationships were established and have already lead to new collaborations that would not have happened without the project taking place.

Despite the wide spread need for quality data modelling, it was apparent that in many organisations and projects, data models are seen as an necessary yet incidental step towards an applied end use rather than as an output in their own right. One way to make more data models openly available would be to require publically funded projects, which involve geoscientific data storage, to produce public versions of the underlying data models produced.

This project has shown that Geological Surveys around the world face similar challenges and require similar data models, the level of knowledge varies greatly but there is a genuine appetite to improve how geoscience data is used, this requires quality data model designs. It is the belief of the project team that there is enough enthusiasm for this topic to justify both:

- a permanent resource for sharing data model designs, ideas and lessons learnt, and

- a workshop or dedicated session at an international geoscience conference focussing on international issues and best practice in the field of geoscience data models.

2. Introduction

1.1 PROJECT MOTIVATION

The British Geological Survey (BGS) has invested well over £1,000,000 since the late 1970s designing and building geosciences databases. This has resulted in a wealth of mature, well documented designs and a reputation for excellence in this area.

Between July 1998 and March 2001 the BGS led a project called "Strategies and Systems for Maximising Data Value - Knowledge and Research (KaR) programme", it was carried out on behalf of the Department for International Development (DFID) and produced a generalised data model for use in geological organisations engaged in geological mapping. The BGS worked with the geological surveys of Botswana, Malawi and Malaysia to provide training on the underlying data model and associated database implementations.

The BGS is regularly contacted by other Geological Survey Organisations (GSOs) and industry requesting copies of database designs. This lead the then head of Informatics, Jeremy Giles and the Data Architect, Martin Nayembil, to recognise the need for an open sharing of geoscience data models amongst a community of GSOs, industry and academics. Various ideas were discussed about how such a community initiative could be developed. Ideas included the setting up of an online repository for data model designs, specialist conferences or workshops to encourage knowledge exchange and targeted technical visits to help develop data modelling expertise in developing countries.

The BGS have actively helped to develop elements of the organisational databases for a number of GSOs, notably GSI (Ireland), GSNI (Northern Ireland) and IGME (Greece) whilst providing designs and data model concepts to many other international organisations and companies.

Through these experiences it became apparent that many geoscience organisations in developing countries lack the essential skills, funds and experience to create the necessary information architecture to effectively manage the information they hold. There is an acute need for better management, storage and presentation of geoscience information. At the same time, there is also a growing need for the standardisation in the way we record such information so that resources and hazards which cross national boundaries can be managed and understood on both sides of a border in the same way. The data model is a key part of successful information management because it provides a centralised description of the meaning and inter-relationships of the information.

A proposal for a two year knowledge exchange project was submitted to the Natural Environment Research Council (NERC) in May 2010. The proposal largely centred on the creation of a community drawn together in the first year of the project by a specially arranged international conference. The proposal paid particular attention to the needs of GSOs operating in developing countries, which is why the conference was due to take place in sub Saharan Africa, with some of the project funds designated for use as travel subsidies for a small number of delegates.

This initial proposal was not funded and based upon the feedback received a second proposal was produced. The revised proposal focussed more on small scale face to face meetings rather than a large scale international conference and benefited from more attention being paid to the sustainability of the community and resources produced. Funding was granted for a 2 year project in February 2011 under the Knowledge Exchange programme, going by the name of OpenGeoscience Data Models, work started in June that year.

PROJECT VISION

"It is our aim to provide open, ready to use database designs that will be available to all on a public website.

We will also build a self sustaining community to develop these and new designs to meet the needs of the wider environmental community."

3. Objectives

The objectives of the OpenGeoscience Data Models project were stated as follows:

- Support NERCs Environmental Monitoring objective of 'Enabling environmental data sets to be applied in new commercial and societal services and products'
- Contribute to the ongoing global improvement in the quality of geoscientific data
- Contribute to the ongoing efforts to ensure water security
- Improve efficiencies within the global community of geoscientific data owners and consumers
- Generate economic opportunities for the BGS and UK plc

In order to meet these objectives the proposal described the creation of a new "Data Models" section within the existing OpenGeoscience website: http://www.bgs.ac.uk/opengeoscience. The main purpose of this section was to describe and give access to the 'Database Design Packages' developed by the BGS. In addition a community of users would be developed who would use and contribute to an ongoing web resource that promotes good database design (& data management) within geoscience.

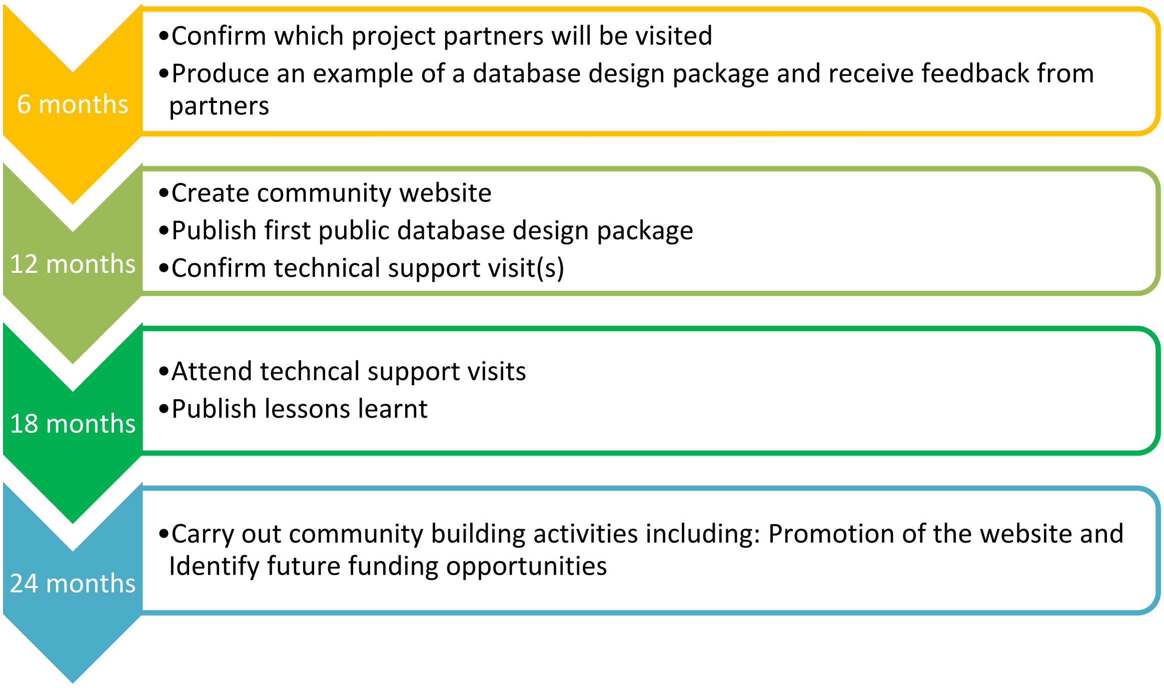
Each Database Design Package would consist of:

- Entity relationship diagrams
- Documentation that was suitable for technical and non-technical audiences
- Programming scripts to create copies of the database designs

It was the belief of the project team and partners that the open sharing of data models in the form of downloadable database designs would lead to better information management within the community. GSOs that lacked data management experience would be able to short cut the learning curve and scientific data end users would benefit from greater access to quality, well described and constrained datasets. Technologically advanced GSOs would also benefit from the peer review aspect of the project.

4. Milestones

The original project plan was to complete within 2 years hitting the following milestones each 6 months.

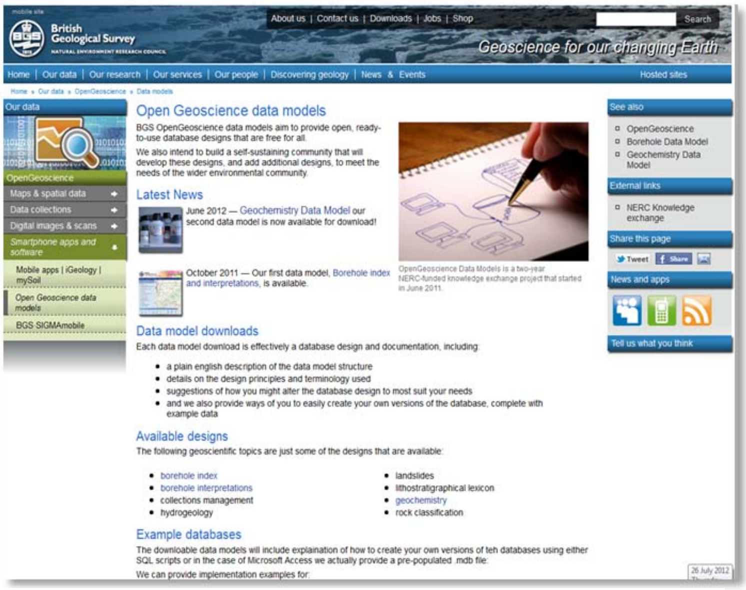


In reality the plan had to evolve as the project got underway due to circumstances beyond the control of the project team, more fully described in the following sections. The main deviations from the initial plan were the extension of the project duration from two to three years, although no extra funding was sought. Digital outputs were produced quicker and with more metadata than initially anticipated, whilst technical visits were more difficult to arrange than expected.

5. Summary of progress made

1.2 HOW IT UNFOLDED

Work started on the first of May 2011, it had been agreed that due to the central role the borehole database plays in the BGS, this should be the first design released. A platform independent data model was produced in the software, ER/Studio (by Embarcadero), this was done by reverse engineering the BGS database and editing the design in accordance with a number of reviews held between the developers and data architect. Once the data model was agreed, the ER/Studio software was used to automatically generate technical specification documentation and implementation scripts for a variety of relational database platforms. The team were very comfortable performing these data modelling steps and the technical outputs were produced relatively quickly. Writing the documentation to make the data model more openly accessible to a wider audience came less naturally and took longer to conclude. A number of BGS staff carried out internal reviews of the documentation and a final draft was circulated amongst project partners with a complimentary copy of a Microsoft Access implementation of the database.



OpenGeoscience Data Models website

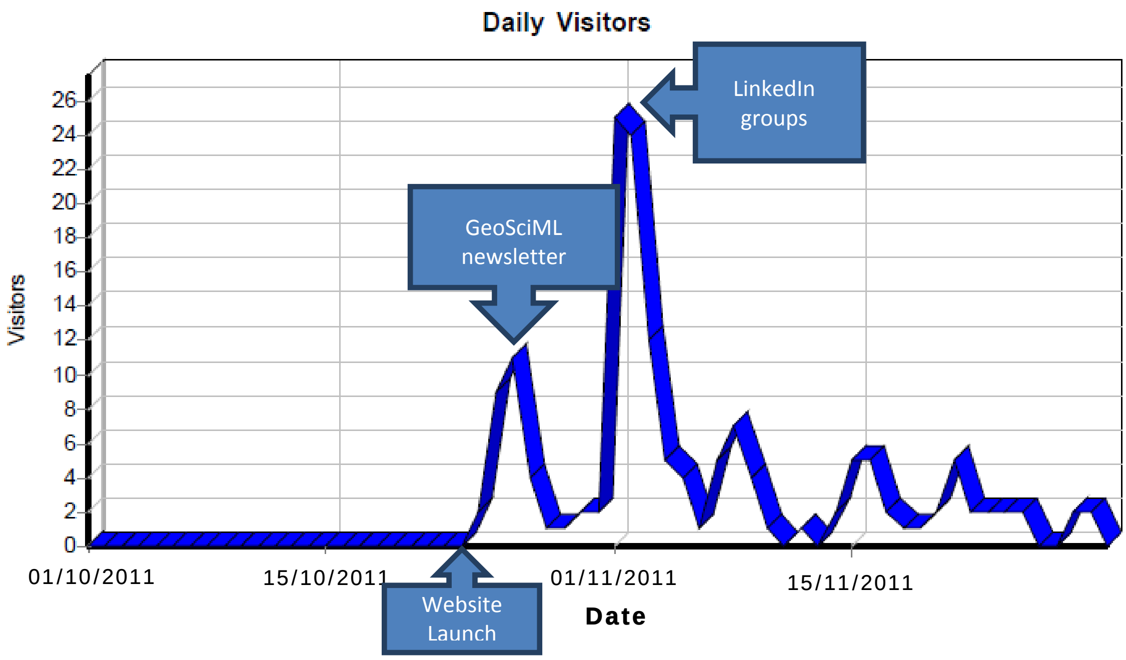
The first website was a relatively simple set of pages which was designed to describe the project objectives and provide access to the newly created downloadable database design package. Feedback was encouraging and comments helped the team to refine the way subsequent documents would be produced.

In parallel with the technical work, **community engagement** took place through presentations at events such as the CGI GeoSciML committee meeting, Edinburgh and arrangements for technical visits to The Minerals Resource Authority (MRA), Papua New Guinea and the Nigerian Geological Survey Agency (NGSA) were made.

Security concerns started to emerge by the end of the first six months and it became apparent that the technical visits were going to be more difficult to arrange than initially thought. Nigeria and Papua New Guinea were project partners since the project planning phase but alternative plans had to be made in light of civil disorder. In the first instance, attempts were made to arrange meetings in alternative locations, this was not ideal, informal conversations were held at conferences attended by project partners but true requirements gathering and implementation activities needed to be performed in situ.

Over the subsequent months attention switched to **development and promotion of the downloadable designs** and take up increased from an average of two or three visitors a day to five or six. The most popular format for data model implementations was Microsoft Access and this appears to be due to the users wanting a quick and simple way to look at the database without executing scripts. This allowed us to stream line the download production for future data models, with only Microsoft Access implementations being produced by default, with other implementation being provided on demand.

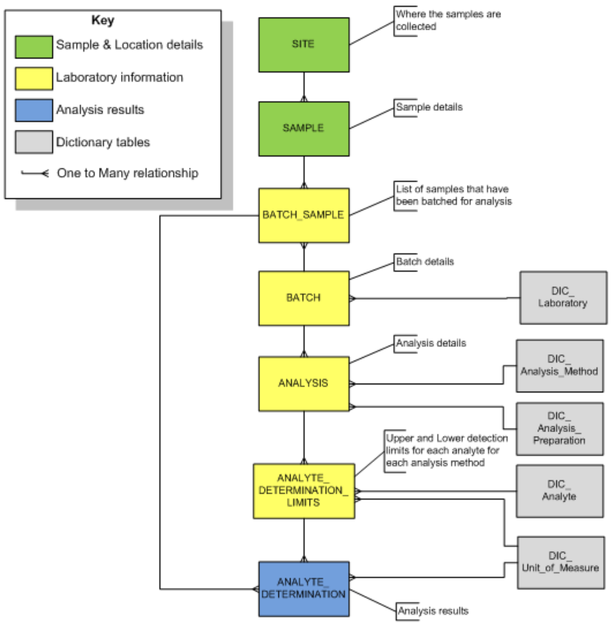
Making use of existing communities and their communication channels lead to significant leaps in visitor numbers, as shown in the following graph. The effect on download numbers after the project was promoted in the GeoSciML community newsletter and the social networking site LinkedIn (through groups such as GEOinformatics and Geocomputing Professionals) was particularly noticeable in these relatively early days.



Download statistics for late 2011

There was some debate regarding the level of intellectual property that should be retained by the BGS/NERC and whether potential users of the outputs should register prior to downloading content. It was decided that the designs would be made available totally free and that they could be used for any subsequent purpose, commercial or other. In order to maximise take up of the downloadable design packages it was decided that registration would not be required despite this preventing the tracking of users.

The second data model to be released was designed for Geochemistry data, this was promoted through LinkedIn, and in particular the creation of a discussion in the International Association of Geochemistry group lead to a notable spike in downloads. A number of questions and suggestions were submitted via the LinkedIn platform and the development team were also contacted directly by UK based commercial laboratories wishing to use the design in a bench testing exercise. In order to satisfy this requirement the scripts for a MySQL implementation were produced.

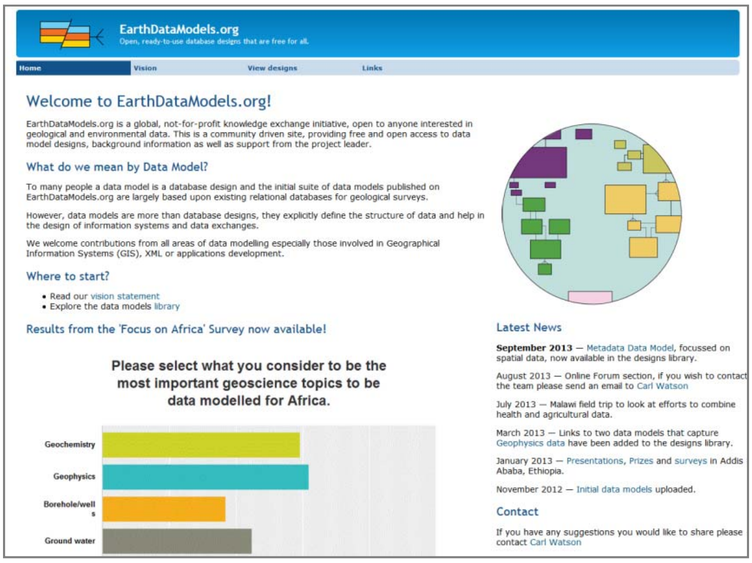


High level Geochemistry data model diagram

Through link ups with other BGS projects, travel costs were shared for the attendance at a number of **international conferences**. A presentation on 'The need for an Open Exchange of Geoscience Data Models' was given by project coordinator, Carl Watson, as an invited speaker at the International Geological Congress (IGC) in Brisbane, 2012. A number of contacts were made at the IGC which lead to the website and contact details being published in the newsletter for the Coordinating Committee for Geoscience Programmes in East and Southeast Asia (CCOP). Informal meetings were held with, amongst others, ground water experts at Queensland University of Technology and project partners MRA of Papua New Guinea. Another notable spike in downloads was observed in the following week.

Despite not being able to visit Papua New Guinea for safety reasons [LL1: R1] contact was regularly made using email and their new database designs were reviewed and recommendations made remotely.

EarthDataModels.org was launched in November 2012, up to this point the online presence for the project was limited to LinkedIn and the BGS OpenGeoscience website so a community website to share data models from the BGS and other organisations was helpful in separating out the project as an initiative independent of the BGS brand [R4]. Primarily populated by BGS designs contacts were made with a number of organisations which lead to the linking up with USGS, GEUS and a few others and the range of data models was significantly increased.



Community website: EarthDataModels.org

The community website included an online forum which initially attracted a reasonable number of users but activity was very limited, most people preferred to contact each other directly via email or phone calls rather than post an enquiry publically. Another problem with the forum was spam and the amount of time required to manage users and filter out unwanted posts [LL2]. In the end the decision was made to close the forum down and focus resources on maintaining direct contacts and producing more data models.

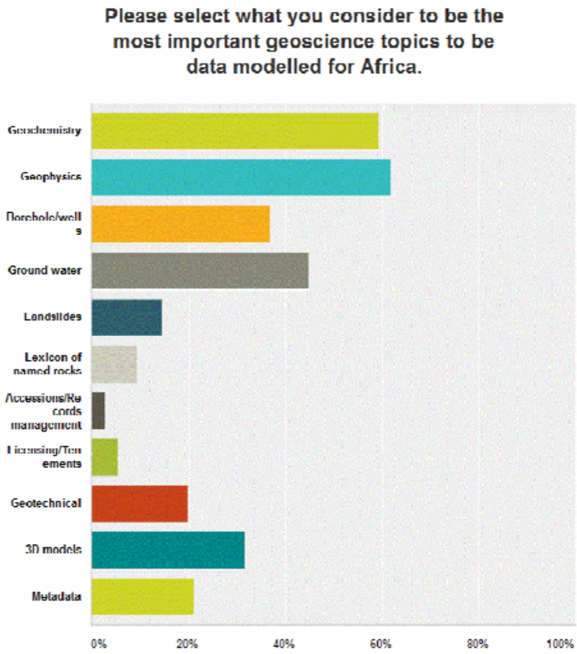
The next significant event took place in January 2013 with two team members attending the Colloquium of African Geology (CAG24), in Addis Ababa, Ethiopia. A presentation was given and a booth was set up for the week from which demonstrations of data models were given and a survey was conducted to assess which subjects were of most value to conference delegates.

The delegates indicated that they wanted geophysics and geochemistry models as their top priorities, with groundwater models coming third, for full details see: http://www.earthdatamodels.org/designs/CAG/surveyResults.html

Data Model demos and questionnaire sessions at CAG24

Over the subsequent months the Lexicon of Names Rocks database design package was released along with links to one of the most popular controlled vocabularies on the Open Geoscience website.

Ad hoc data modelling enquiries were received on a regular basis and dealt with by the team, these included contacts from an GSO in Bhutan who were analysing options for digitising their spatial data holdings, the Swedish Geological Survey asking for designs to help with the redesign of their borehole database and a South African software developer interested in producing a Laboratory Information Management System (LIMS) based upon the Geochemistry design published by the project.



Results for one of the questions asked of CAG24 delegates

IGS (International Geoscience Services) Ltd, a UK based company providing geoscientific and geodata services to the global market, started the IGS Geodata project during the first year of the OpenGeoscience Data Models project. IGS Geodata is a dedicated effort to create a geodata management and enrichment system. The project focuses on regions in which mineral exploration industry has a considerable potential to grow but lack central data repository and geodata has to be acquired separately and then combined.

From an email dated 06/11/2014, Sławomir Wójcik said: *"Members of the Geodata project team has met Mr Carl Watson as he offered help and advice regarding his previous experiences in a similar field. He described BGS efforts in this area in detail and guided the team to find out more about BGS Linked Data project. A fair amount of time has been spent investigating BGS achievements and the Earth Data Models database designs. We followed the advice of Mr. Watson to take a closer look at OpenGeoScience data models and standards proposed by CGI. Mr. Watson talked about GeoSciML, an industry standard language that is used to describe geodata, being an extension of well-known XML standard.*

As a result of Mr Watson's involvement, Geodata project benefited by adopting several industry standards we would have to spend a considerable amount of time and effort to find. A suitable, well tested ontology provided by CGI has been selected and adapted to Geodata project needs. GeoSciML is also considered as a part of data interchangeability module in the future. BGS Linked Data website and endpoint also served as an inspiration for how to open the databases in a succinct and efficient way."

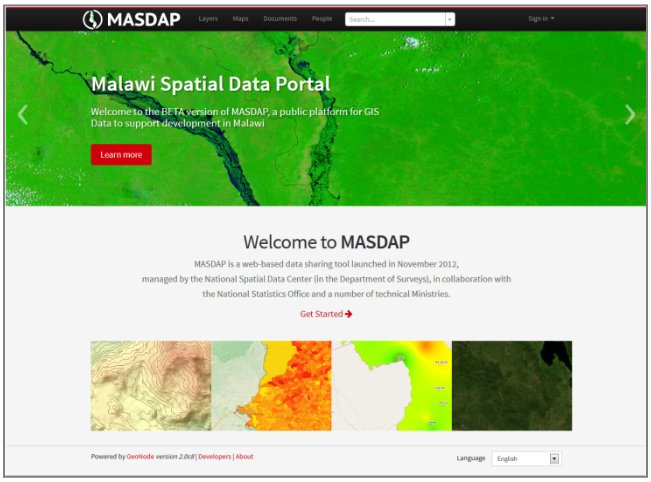
Malawi was the destination for a ten day technical visit in June 2013, the project-coordinator hosted and attended a number of meetings to investigate data modelling capacity within departments responsible for spatial data management, identify potential collaboration opportunities and promote the free to use data models published at www.earthdatamodels.org. The visit was primarily based in Lilongwe and Blantyre, with appointments at institutions such as the Ministry of Health, Ministry of Agriculture and Surveys Department.

There are a large number of projects run by government, commercial organisations and Non-Government Organisations (NGOs) in the country that involve the capture and storage of spatial data, methods vary from organisation to organisation. There is a high regard for GIS expertise in the country and plenty of well trained and keen practitioners, there is also a significant number of researchers and decision makers who are interested in finding out what spatial datasets exist within the country. Unfortunately there appear to be a large number of datasets that could have national significance but remain hidden on individual laptops, often held in spreadsheets and GIS files that are well understood by the authors but poorly documented, raising the risk of data being lost or misunderstood by future potential users.

The June visit lead to a subsequent visit in February 2014 which was more technical in nature and was very much focussed on supporting the work being carried out by the Department of Surveys.

One shining example of how spatial data management could be improved in Malawi, and perhaps other low income countries, is http://www.MASDAP.mw. This system provides an online portal for the upload, search and download of spatial datasets, its development is funded by the World Bank programme for Disaster Risk Management.

The Department of Surveys are in the process of creating a National Spatial Data Centre for the country, they will host and administer the MASDAP system and are in the process of developing national standards for the capture and storage of spatial data.



Malawi spatial data portal: MASDAP

The MASDAP system is a powerful tool for capturing and storing spatial data, however, it is only possible to capture metadata when a user is uploading a dataset. Many data owners are willing to tell others what data they possess but strongly resist ‘giving it away’. We came up with a plan to encourage the capture and use of spatial metadata through a series of changes to the existing system and by arranging data management and data modelling training in the UK and Malawi. The BGS data model for spatial metadata and information workflows were presented and used to inform discussions on future developments.

Over the two visits we met a lot of welcoming people, in particular Alice Gwedeza at the Surveys Department and Allan Chilimba at the Ministry of Agriculture who treated us with great enthusiasm and hospitality.

Email dated 06/03/14 from Alice Gwedeza: *“I am writing this with great pleasure to thank you for the enthusiasm that you have for assisting Malawi to come up with a better database infrastructure specifically in developing the metadata for Malawi. Further I want to thank you for your efforts in not just seeing something but to come and discuss with us in order to map the way forward. Honestly, as a mapping/GIS profession I am dying to see a well organised, up to date and comprehensive spatial data Infrastructure to avoid duplication of efforts whereby people collect data and do not share as a result the data for the same area can be collected several times by different users.*

We are very eager to work with assistance from the British Geological Survey. Sometimes here we have the knowledge but we are frustrated by lack of resources, as a result a lot of resources are wasted. With the metadata in place we will be assisted in knowing which data sets exist.

We are looking forward to work with you to achieve the production of metadata."

In order to cement the relationships with the Malawi partners a successful bid was submitted to the Commonwealth Scholarship scheme to fund a member of the Department of Surveys to travel to the UK for data modelling and information management training.

The final months of the project involved significant effort to forge new connections with European GSOs and others who would be interested in collaborating on data modelling and data management issues after the initial knowledge exchange project ceased. In particular links with the Geological Survey of Denmark and Greenland (GEUS) through the COST Sub-Urban network workshops lead to follow up knowledge exchange meetings between the BGS data architect and GEUS head of information systems and new collaborative opportunities continue to be discussed. A presentation was made at the commercial and public sector GSPEC community in Glasgow 2014, to promote centralised and well controlled data management to support geotechnical engineering activities. These efforts have lead to several members of the project team being invited to take part in new international projects within the field of geoscience data modelling.

6. Challenges encountered

1.2.1 Staff turnover

During the lifetime of the project there was significant disruption caused by the departure of several developers and transfer of the principle investigator role on two separate occasions. The enthusiasm of the remaining team members and the fact the head of Data Science, Garry Baker, kept abreast of developments throughout the project helped to ensure the project remained largely on track.

1.2.2 Security issues

As briefly mentioned in the previous section arrangements were made in the project planning phase to visit Papua New Guinea and Nigeria but due to events in both countries which resulted in the FCO security risk assessment being raised to High, it was decided by the BGS Senior Management that this visits could not be authorised.

1.2.3 Community creation and engagement

Despite there being significant enthusiasm for sharing knowledge on data models, associated IT architecture and business workflows amongst a wide range of contacts there was no pre-existing community of geoscientific, or more general environmental, data modellers. The aim to create and maintain such a community in such a short time whilst also producing a significant number of outputs has proved to have been very ambitious. One possible reason for the absence of such a community might be that many data modelling practitioners only actively engage in research and development when there is a specific end goal, often in response to the need for a new relational database or because an existing one requires re-engineering.

If an initiative such as the Research Data Alliance (RDA) had existed at the time this project was being planned it may have proved a useful vehicle for such a venture. Another approach could have been to contribute data modelling expertise to multi-national projects that involve the creation of geoscientific databases. The data modelling activities required in the production of databases for publically funded research could and, we would argue, should involve adequate resources to produce publically available data models [R2].

7. Results & Highlights

1.3 ONLINE, DOWNLOADABLE DATA MODEL PACKAGES

As briefly described in the objectives section the project produced a number of downloadable database design packages. Initially these were referred to as data model downloads packages but it became apparent that the vast majority of potential users were non-database experts and therefore did not necessarily understand the specialised field of data modelling, but they did understand that databases required an underlying design.

The downloads were made up of relational data models in the form of entity relations (ER) diagrams, plain English documents (available in MS Word and PDF to maximise access), implementation scripts, example data and links to associated dictionary data sources.

The BGS staff on the project produced four full database design packages and a further three are in progress, which, will be released regardless of the knowledge exchange project formally coming to an end.

The BGS produced downloads were complemented by a number of designs made available by project partners and associated organisations, these were hosted on the information providers website and links provided on EarthDataModels.org model library.

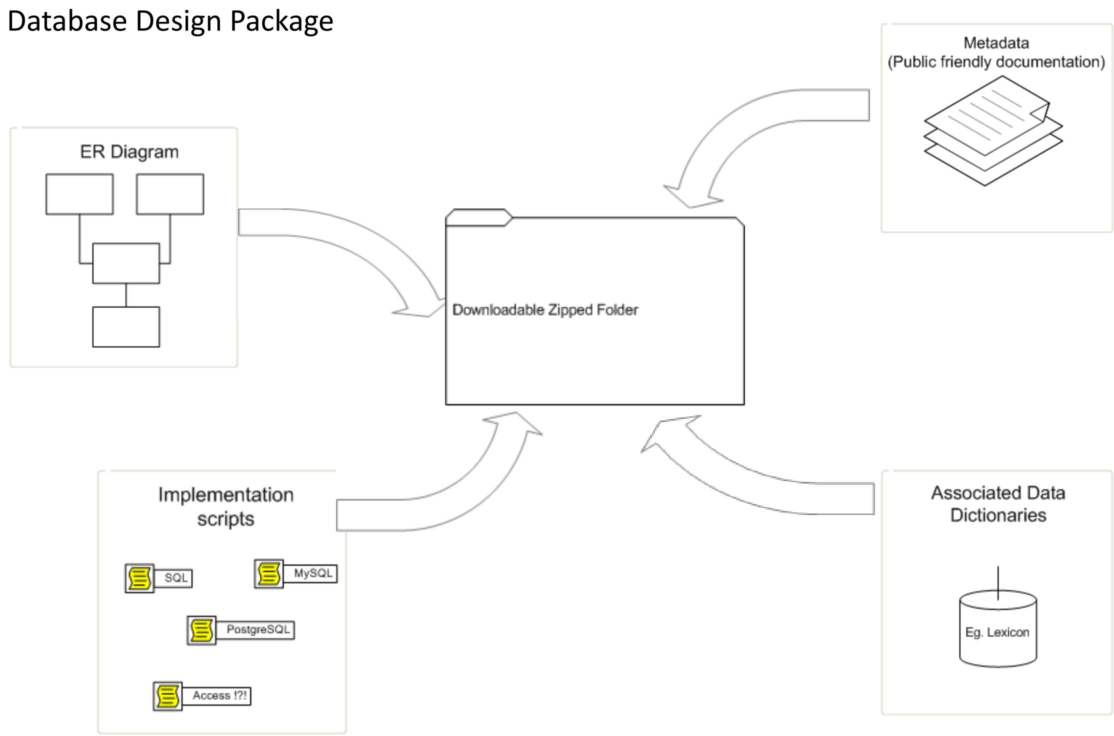


Diagram to show the components which make up a downloadable database design package

1.3.1 Table showing data model downloads created specifically for the project:

Topic	Download items	Status
Borehole	• Logical data model documents • Microsoft Access implementation and technical documentation • MySQL scripts and technical documentation • Oracle scripts and technical documentation • PostgreSQL scripts and technical documentation • SQLServer scripts and technical documentation	Live
Geochemistry	• Logical data model documents • Microsoft Access implementation and technical documentation MySQL scripts were created and distributed directly to some users	Live
Lexicon	• Logical data model documents • Microsoft Access implementation and technical documentation	Live

Metadata	• Logical data model documents • Microsoft Access implementation and technical documentation	Live
Landslides	Draft data model documentation and Entity Relationship model created.	Unfinished
3D geological model objects	Entity Relationship model created.	Commitment made by the BGS to release design post peer review paper publication
Coal Analysis	Draft data model documentation and Entity Relationship model created.	Commitment made by BGS Wales to release by the end of 2014/2015 financial year

1.3.2 Table showing data models made available by community members and associates:

Topic	Provider
Borehole	Geological Survey of Denmark and Greenland (GEUS)
Geochemistry	U.S. Geological Survey (USGS)
Geological Mapping	The North American Geological Data Model (NADM)
Geological Mapping	GeoSciML
Geological Mapping	New Mexico Bureau of Geology and Mineral Resources
Geophysics	Geological Survey of Denmark and Greenland (GEUS)
Geophysics	XODB by Xstract mining consultants, Australia
Oceanographic	British Oceanographic Data Centre
Water Quality	U.S. Geological Survey (USGS)

1.4 DOWNLOAD STATS FOR ONLINE DATA MODEL PACKAGES

Year	Total Hits (downloads)	Unique Visitors	Most poplar download
2011	655	188	Boreholes
2012	2,531	620	Boreholes
2013	3,618	578	Geochemistry

* Note: these do not include the direct delivery of data models via email and in person and the most popular data model to be delivered directly to contacts described spatial metadata.

1.5 CONFERENCES

Conference	Presentation title
IGC Brisbane 2012	The need for an Open Exchange of Geoscience Data Models
CAG Addis Ababa 2013	Sharing the most valuable database designs in African geology
IMAG Madrid September 2013	Control the system, free the data, feed the science

1.6 TECHNICAL VISITS, WORKSHOPS AND MEETINGS

Location	Objective	Outcomes
Malawi - June 2013	Assess data modelling capability and requirements	Plan of action for return visit(s) and new contacts made
Malawi - January 2014	Detailed requirements gathering with the Department of Surveys	Spatial metadata model handed over, plan of action agreed with DoS and World Bank to implement a new spatial metadata system for Malawi
Denmark - February 2014	Share understanding of data models amongst GSOs, City planners and water companies	Follow up meeting between Data Architects set up between BGS and GEUS. Invite to join pan European knowledge exchange project Sub-Urban, specialising in Data Management.
Denmark - April 2014	Knowledge Exchange on data models, systems architecture and related information management issues.	Agreement to collaborate in the future on various data models including geotechnical data and 3D model storage.
UK, IGS Face to face meetings held: 12/04/2012 21/12/2012 17/01/2013 25/10/2013 Email support: 02/04/2014 15/04/2014	Identify data models which could support the commercial aspirations of the newly formed IGS company	IGS have incorporated lessons learnt from BGS development of databases and use of international standards to develop the new commercial produce IGS Geodata.

8. Lesson Learnt

LL1: When working with developing countries there is a real danger that the security situation can change rapidly and travel plans which seemed reasonable in the planning phase of a long running-project may become impossible within months. Consider alternatives and have, at least high level, back up plans in place.

LL2: Creating and maintain a new community discussion forum takes a lot of time and effort from a setup and administration point of view, if such a facility is justified then it is possibly better to create a group within an existing social media platform, such as LinkedIn or Google Groups.

LL3: Building a community without making funds available to all parties for the travel and expenses required to carry out the necessary face to face contacts seriously limits the effectiveness of the community.

9. Conclusions & Recommendations

1.7 CONCLUSIONS:

C1: Geological Surveys around the world face similar challenges and produce similar data models

C2: Face to face communication produces better outputs and stronger relationships/networks than remote methods

C3: Many of the experiences developed within organisations such as the BGS are directly relevant to organisations responsible for the management of spatial data, not just those involved in the geosciences.

C4: As with C3, above, there are a few organisations, such as BGS, who could share a wealth of experience with regards to producing commercial outputs by deriving digital products from well structured data.

C5: The project team delivered on each of the stated objectives, however, efforts to address the following objective were limited: “Contribute to the ongoing efforts to ensure water security”. Despite providing links to relevant existing public data models on EarthDataModels.org the project team did not contribute a new design, this was due to the priorities identified by our project partners during the mid to latter phases of the project.

1.8 RECOMMENDATIONS:

R1: Future project plans for similar initiatives should include contingency for disruption to travel plans due to security concerns, especially for countries which are identified through the UK Foreign and Commonwealth Office Travel advice service as being prone to high levels of crime or potential for terrorism threats, even if the situation appears calm during the project planning phase.

R2: Future publically funded projects which involve the creation of geoscientific databases should include adequate resources to produce public versions of the underlying data models used or produced.

R3: There are enough organisations actively engaged in geoscience data modelling to justify a workshop or dedicated session at an international geoscience conference.

R4: Use the most appropriate branding for dissemination of public knowledge and if necessary use more than one communication channel, tailoring each to the intended audience. The EarthDataModels.org website was deliberately ‘Open’ and community branded, rather than BGS branded. This worked well when attempting to encourage technically sophisticated project partners and others to contribute material as these contributors tended to have their own strong brands which they were happy to associate with a project initiative rather than appear to be part of a another ‘rival’ brand. However, there were certain other organisations that were encouraged to take part because the project was lead by a well respected organisation, these tended to be the less technically developed institutions such as the GSOs in sub-Saharan Africa.

2 Appendix

This appendix contains a series of screenshots and diagrams that relate to the work carried out and outputs generated throughout the project.

2.1 WEBSITE SCREENSHOTS

Open Geoscience Data Models

BGS Open Geoscience Data Models aim to provide open, ready-to-use database designs that are free for all.

We also intend to build a self sustaining community that will develop these designs, and add additional designs, to meet the needs of the wider environmental community.

You can influence which topics we develop by filling in our short online survey

Photo copyright: iStockphoto.
Open Geoscience Data Models is a two-year NERC-funded knowledge exchange project that started in June 2011.

Latest News

Borehole Data Model
October 2011 — Our first data model is now available for download!
It contains the design and documentation required to create a database for holding a Borehole Index and Geological Interpretations, see Borehole Data Model

What will we deliver?

The BGS and its project partners will deliver the following:

- downloadable, topic specific 'database design packages', containing:
  - fully documented data models
  - implementation scripts in a variety of formats
  - related data dictionaries
- technical support, through:
  - documentation
  - online tools, such as forums and wikis
  - site visits to help some users implement designs
- a plan to promote and continue the Open Geoscience Data Models initiative after the initial project life of two years

The original website, part of the BGS OpenGeoscience resource: http://www.bgs.ac.uk/opengeoscience/home.html

http://www.bgs.ac.uk/services/dataModels/home.html

OpenGeoscience Data Models | Borehole index & interpretations

The Borehole Data Model is our first database design package to be released.

This data model is designed to represent an index of boreholes linked to their geological interpretations and associated metadata.

The design also makes use of corporate dictionaries/controlled vocabularies such as the BGS Lexicon.

The BGS borehole drilling team (Dando rig) in action.

Why use a borehole data model?

The BGS have been using databases to manage our borehole records for over 15 years, the data models behind those systems have continually evolved over that time to provide greater efficiencies and functionality.

Here are a few of the benefits our current database provides:

- single central location for holding millions of borehole records
- it is possible to receive large volumes of new boreholes and quickly make that information available for use by geologists across the organisation
- we can extract the locations for all our borehole records and view them in a GIS
- the design of the database helps standardise information from a variety of sources
  - controlled vocabularies and logical contraints make the information more re-usable and discoverable
  - data can be formatted and extracted for a range of uses
  - clearly defined tables and columns remove the likelihood of ambiguity within the dataset
- 'Single version of the truth' — possibly the most valuable benefit in using a centralised database to manage geoscience information

Downloads

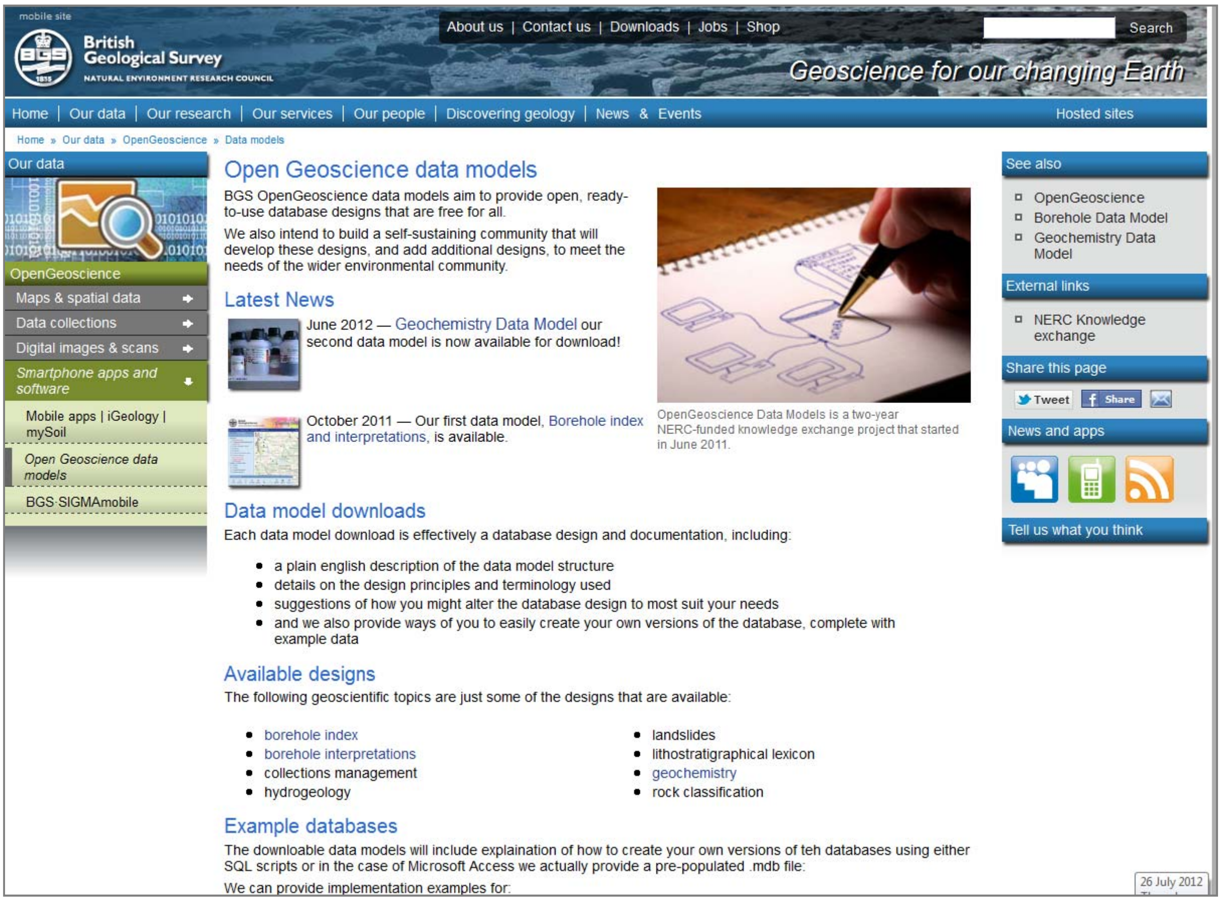
Select the database platform you are most interested in or simply select the Logical Data Model only option from the following list:

- Logical Data Model only download

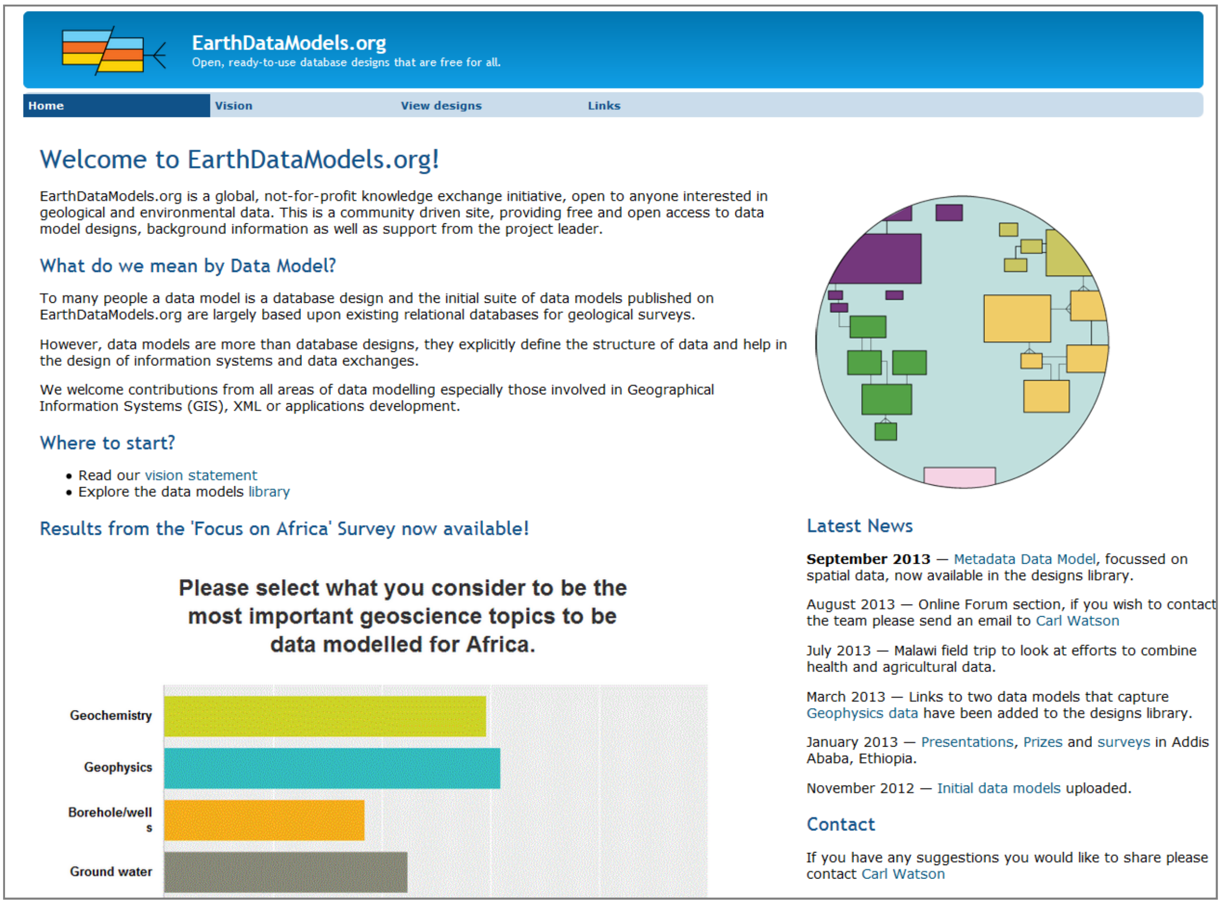
The implementation scripts are available for the following relational database platforms:

- Microsoft Access download
- MySQL download
- Oracle download
- PostgreSQL download
- SQLServer download

The Borehole Data Model webpage

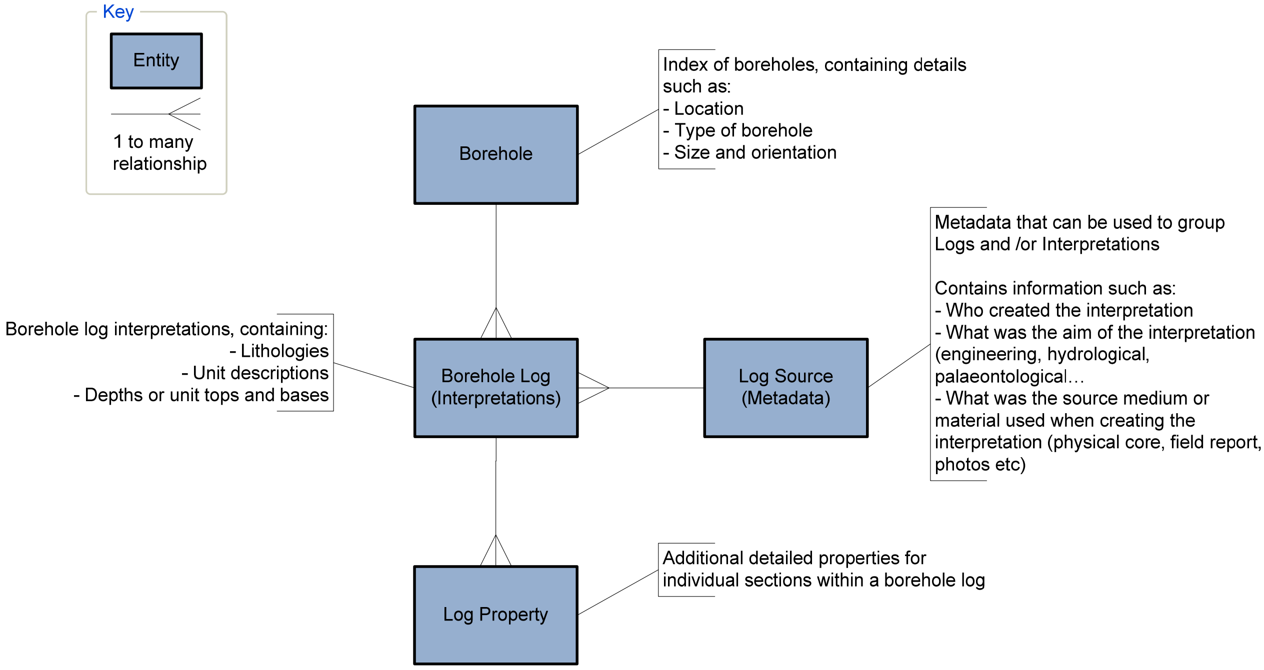


The original website as it was in July 2012: http://www.bgs.ac.uk/services/dataModels/home.html

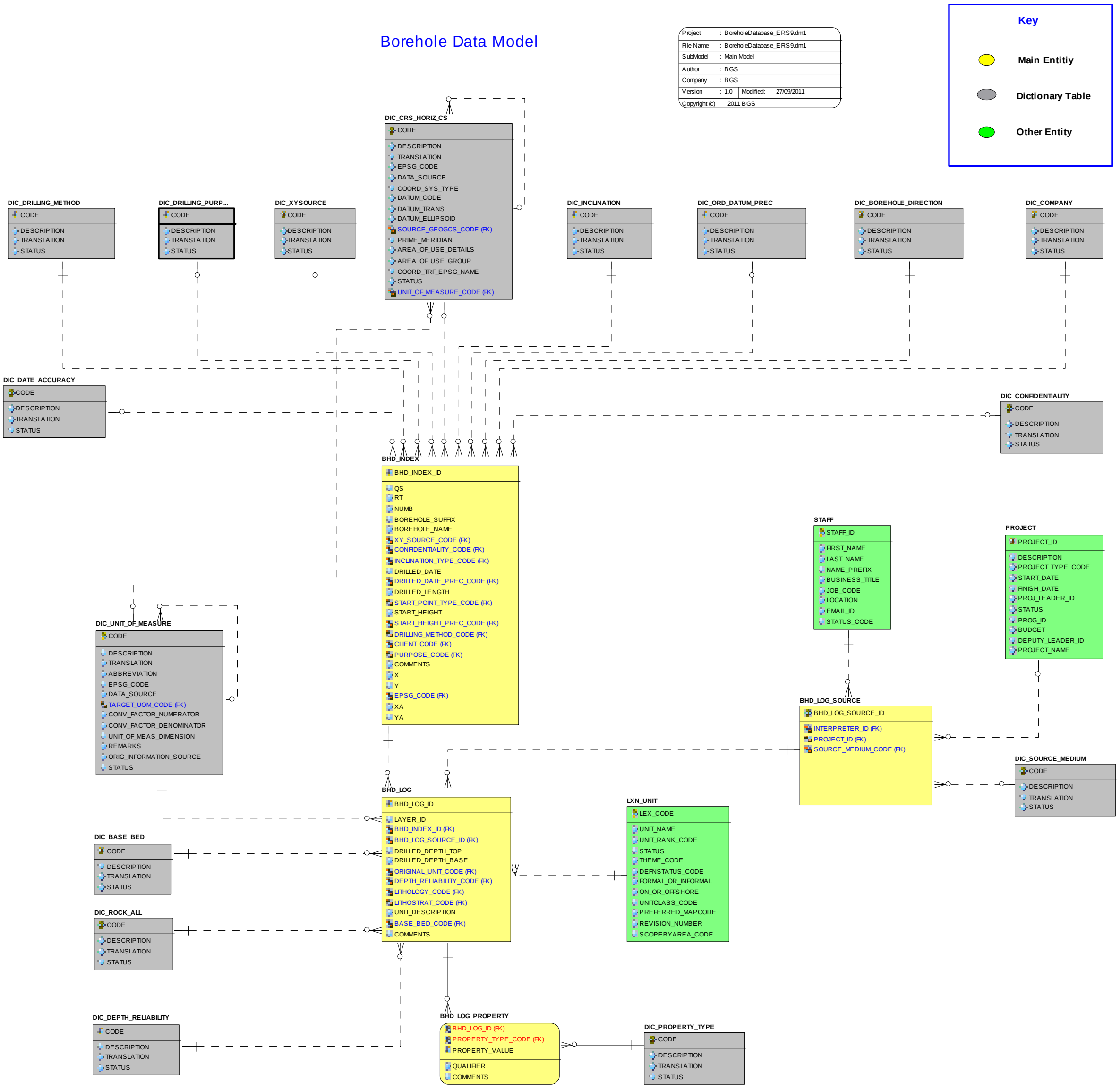


The homepage for the community branded (rather than BGS) website: http://www.earthdatamodels.org

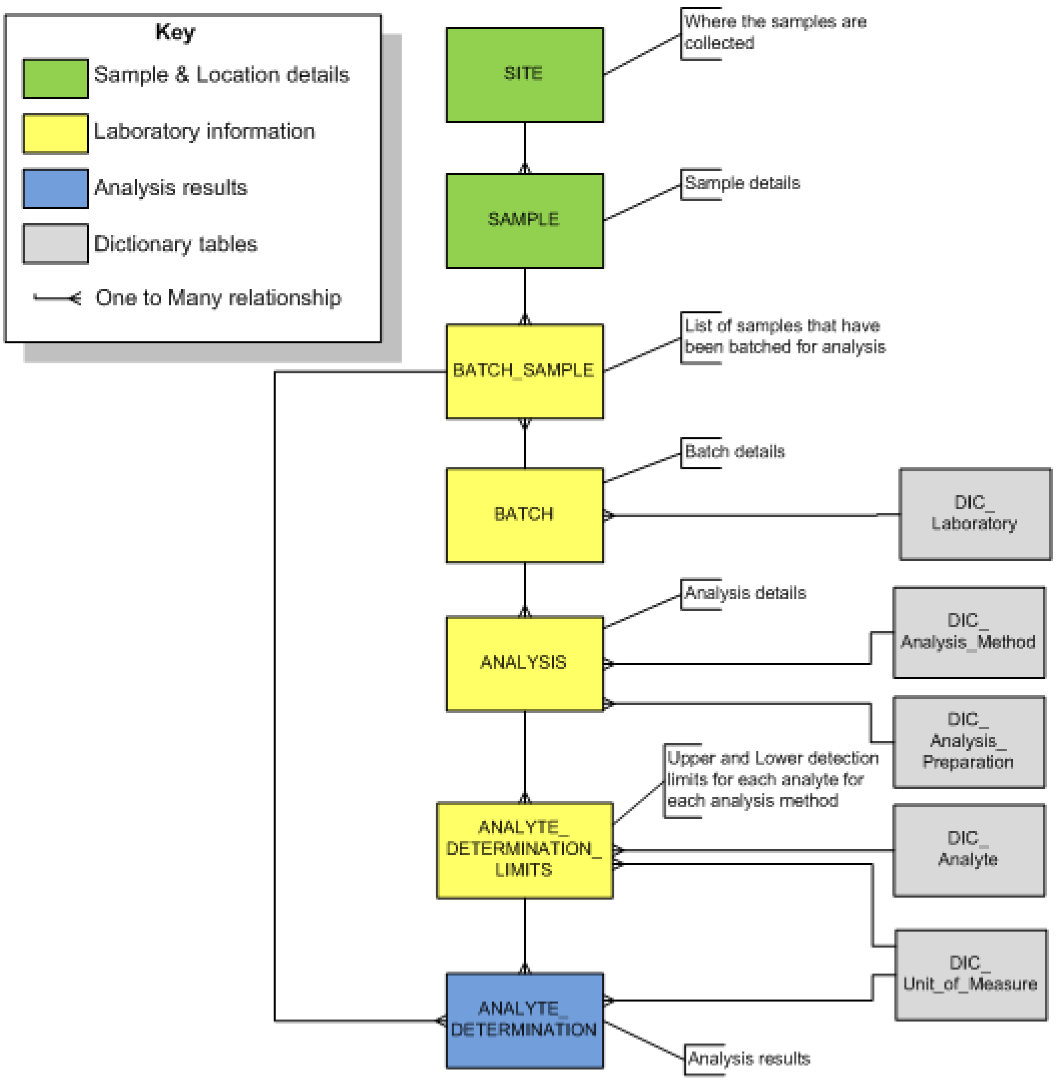
2.2 SELECTED DATA MODEL DIAGRAMS



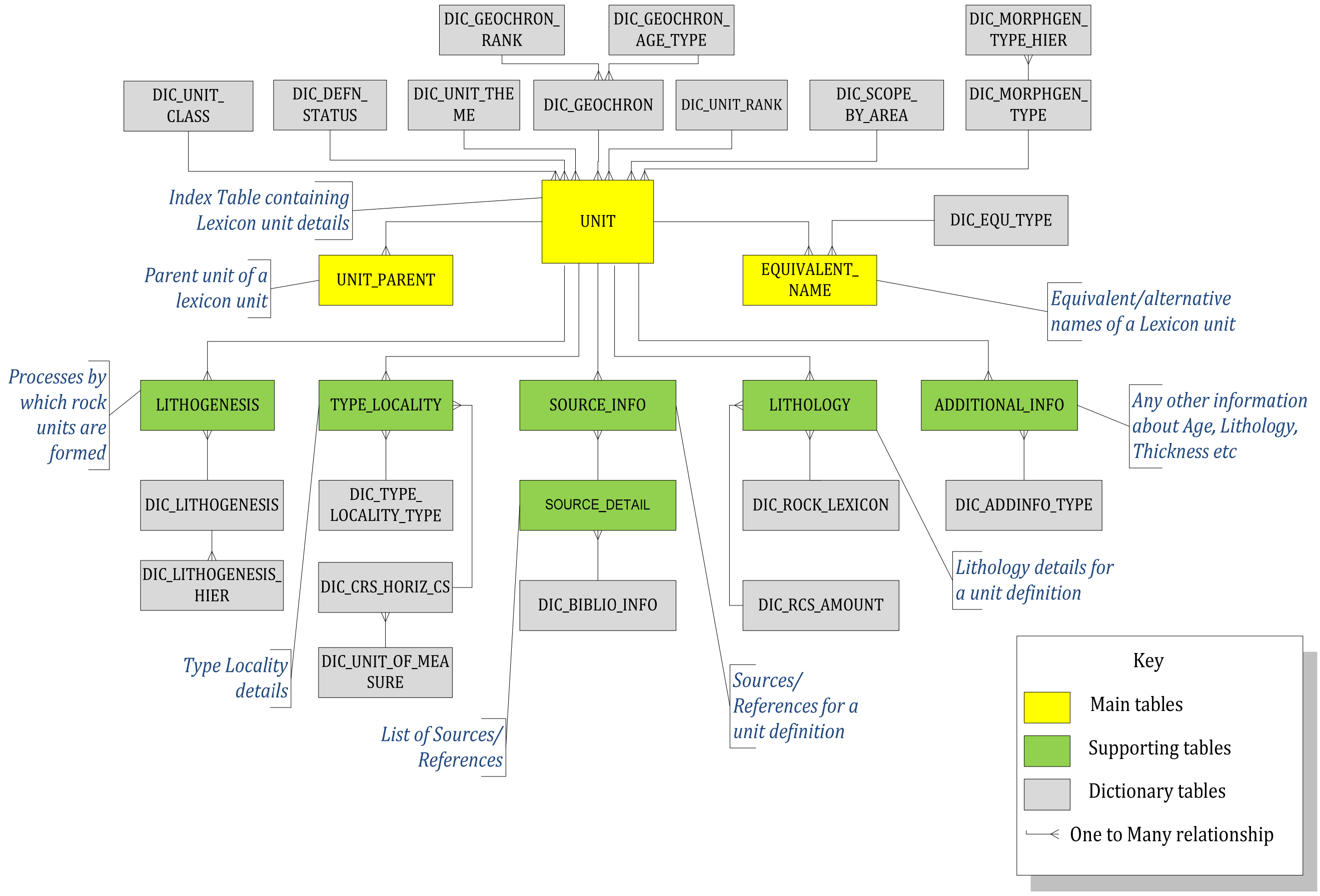
High level entity relationship diagram for the borehole data model



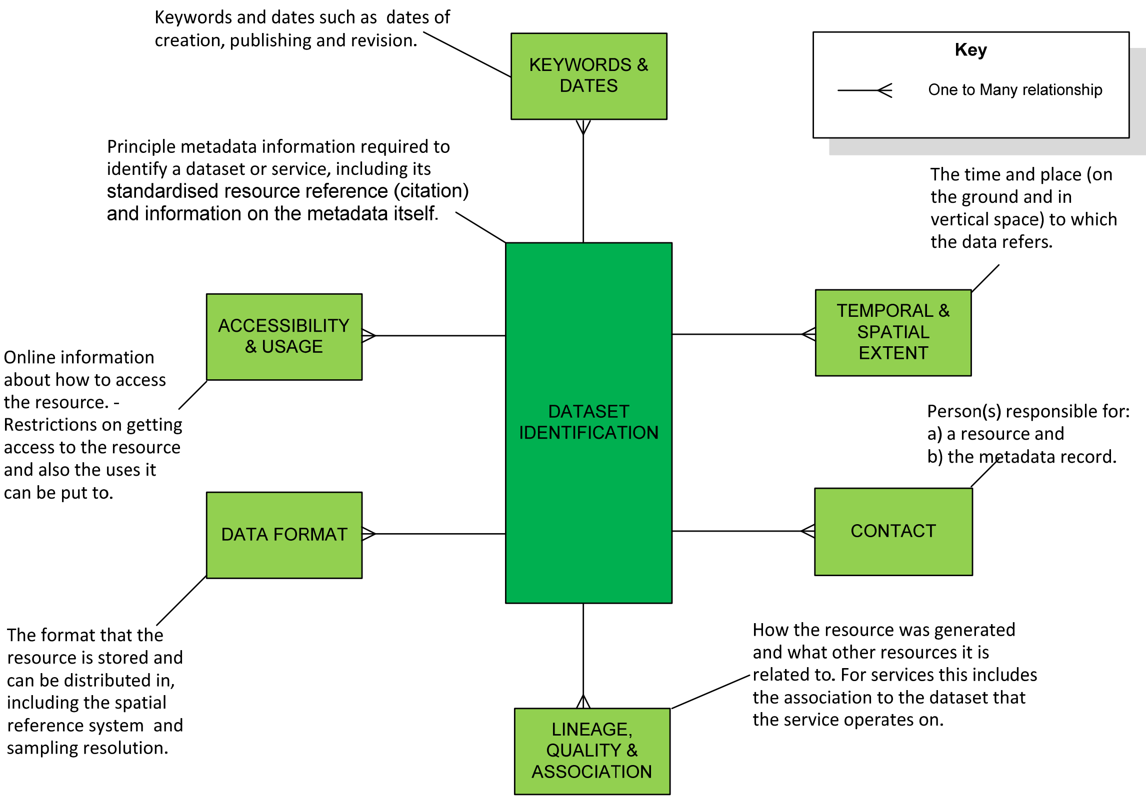
Full entity relationship model diagram for the Borehole data model



High level entity relationship model diagram for the Geochemistry data model

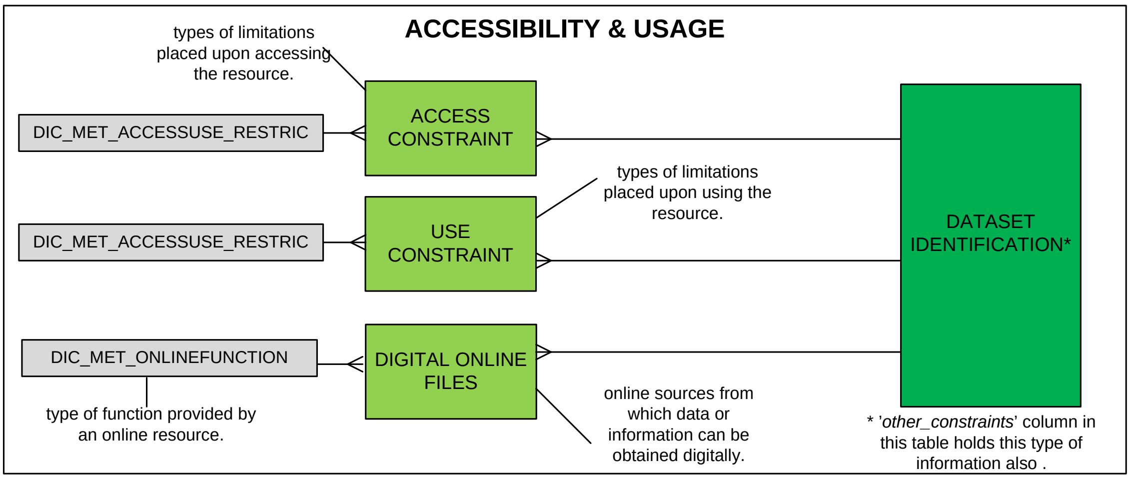


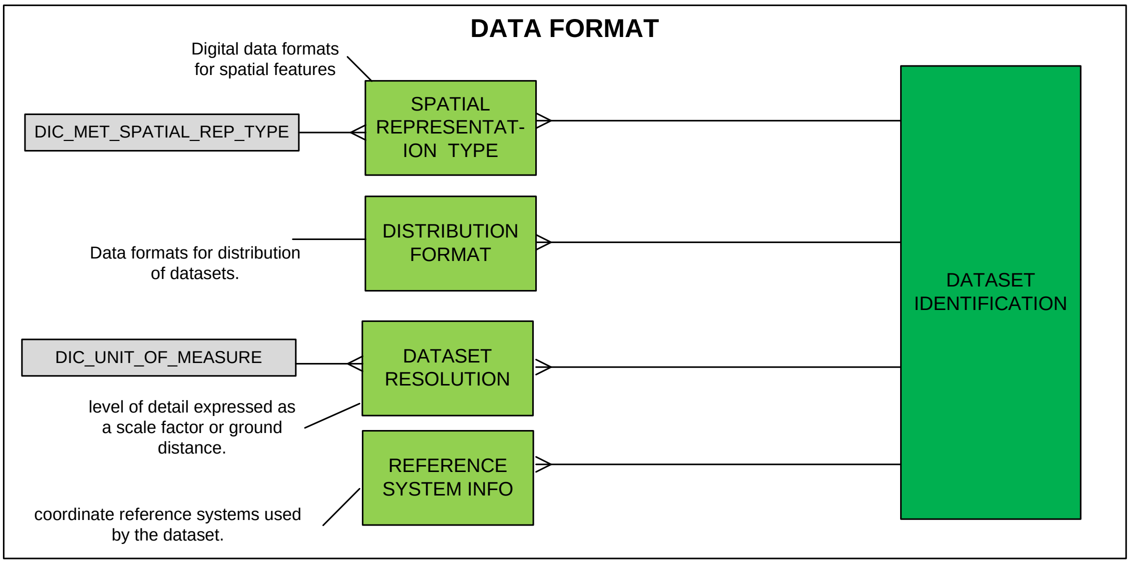
High level entity relationship model diagram for the Lexicon of Named Rocks data model

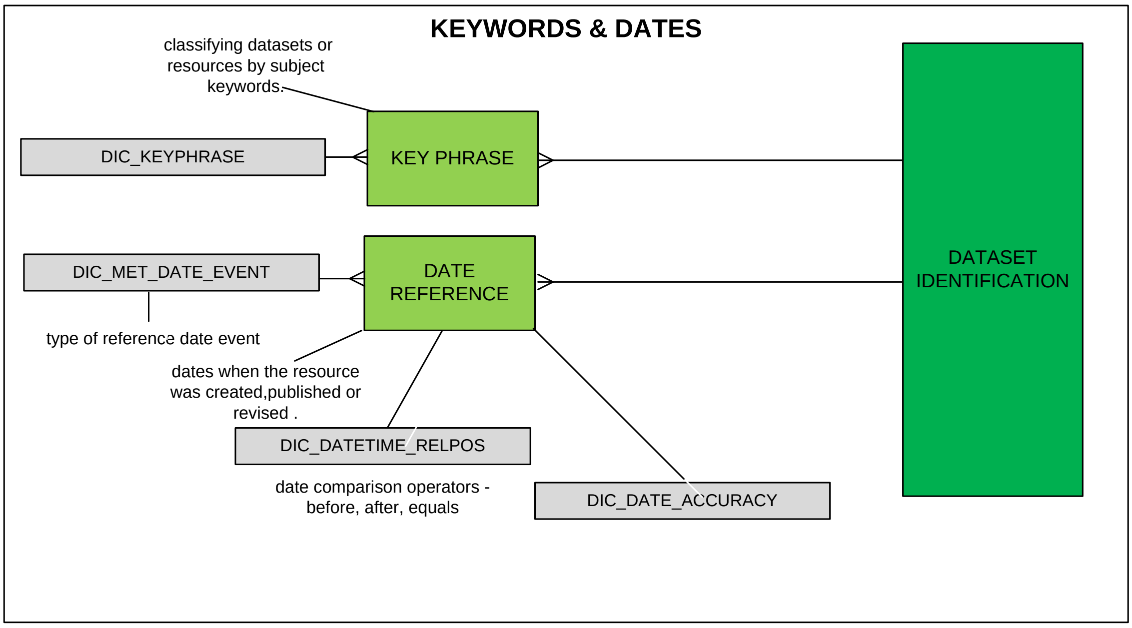


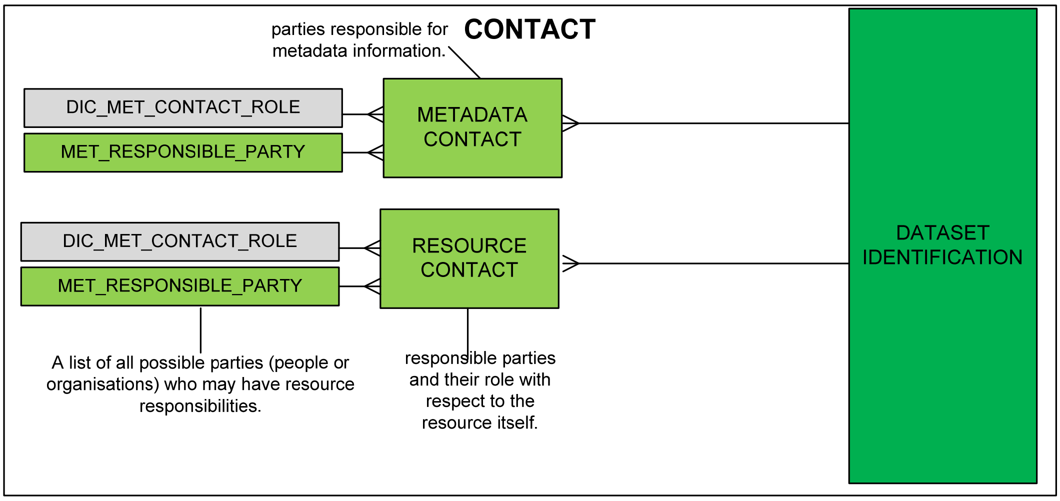
High level diagram for the spatial metadata data model.

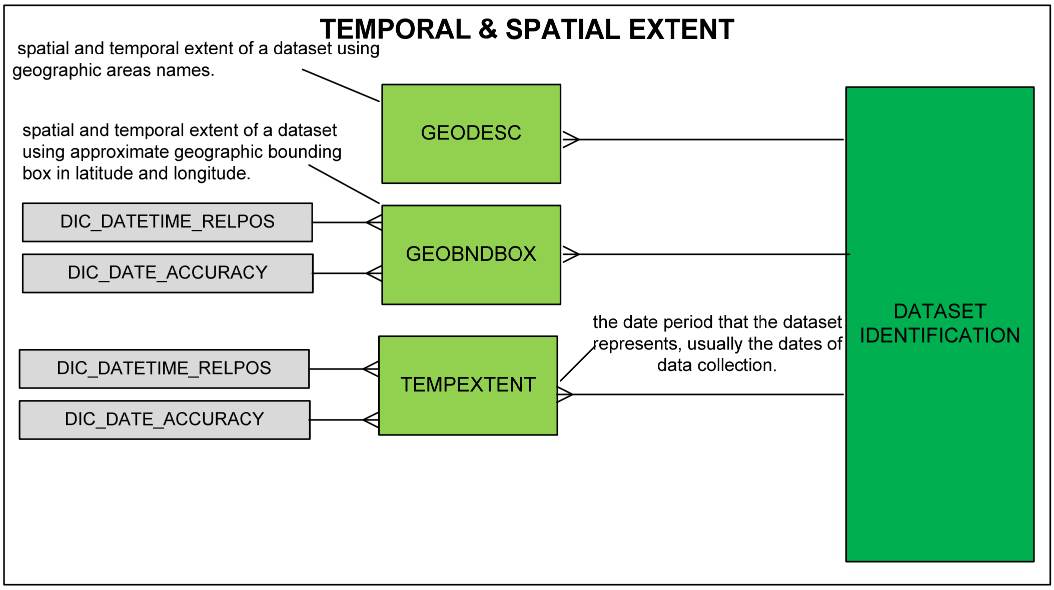
Each of the child tables (light green) in the high level diagram were further explained by additional figures showing the logical structure of the data model, as shown over the next two pages.

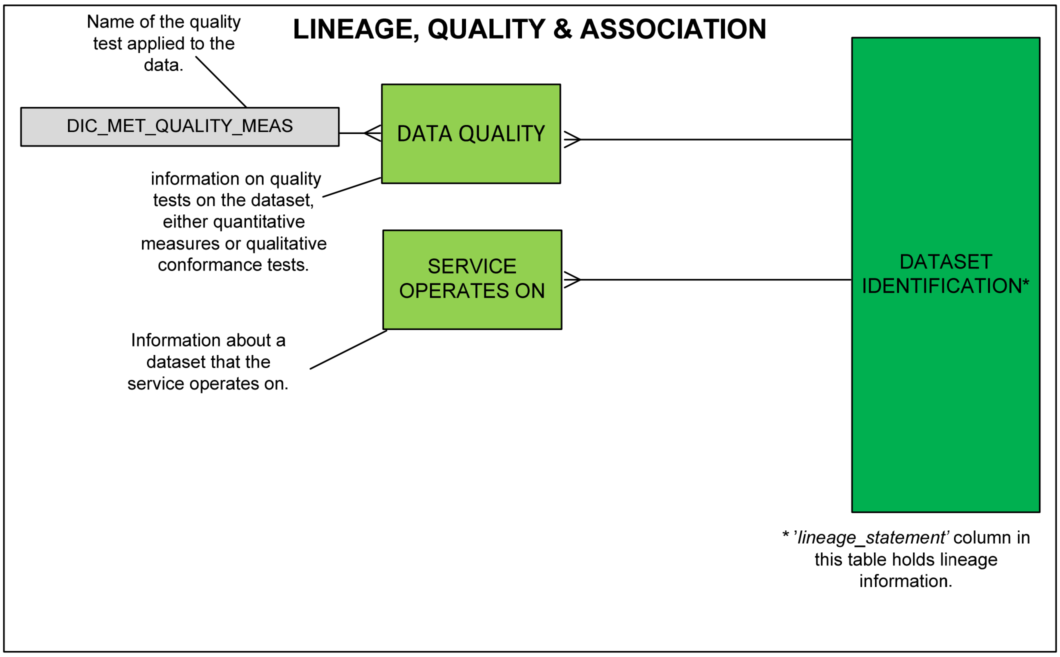

ACCESSIBILITY & USAGE
types of limitations placed upon accessing the resource.
DIC_MET_ACCESSUSE_RESTRIC
ACCESS CONSTRAINT
DIC_MET_ACCESSUSE_RESTRIC
USE CONSTRAINT
types of limitations placed upon using the resource.
DATASET IDENTIFICATION*
DIC_MET_ONLINEFUNCTION
DIGITAL ONLINE FILES
type of function provided by an online resource.
online sources from which data or information can be obtained digitally.
* 'other_constraints' column in this table holds this type of information also .


DATA FORMAT
Digital data formats for spatial features
DIC_MET_SPATIAL_REP_TYPE
SPATIAL REPRESENTAT-ION TYPE
DISTRIBUTION FORMAT
Data formats for distribution of datasets.
DIC_UNIT_OF_MEASURE
DATASET RESOLUTION
level of detail expressed as a scale factor or ground distance.
REFERENCE SYSTEM INFO
coordinate reference systems used by the dataset.
DATASET IDENTIFICATION


KEYWORDS & DATES
classifying datasets or resources by subject keywords.
DIC_KEYPHRASE
KEY PHRASE
DIC_MET_DATE_EVENT
DATE REFERENCE
type of reference date event
dates when the resource was created,published or revised .
DIC_DATETIME_RELPOS
date comparison operators - before, after, equals
DIC_DATE_ACCURACY
DATASET IDENTIFICATION


CONTACT
parties responsible for metadata information.
DIC_MET_CONTACT_ROLE
MET_RESPONSIBLE_PARTY
METADATA CONTACT
DIC_MET_CONTACT_ROLE
MET_RESPONSIBLE_PARTY
RESOURCE CONTACT
DATASET IDENTIFICATION
A list of all possible parties (people or organisations) who may have resource responsibilities.
responsible parties and their role with respect to the resource itself.


TEMPORAL & SPATIAL EXTENT
spatial and temporal extent of a dataset using geographic areas names.
GEODESC
spatial and temporal extent of a dataset using approximate geographic bounding box in latitude and longitude.
DIC_DATETIME_RELPOS
DIC_DATE_ACCURACY
GEOBNDBOX
the date period that the dataset represents, usually the dates of data collection.
DIC_DATETIME_RELPOS
DIC_DATE_ACCURACY
TEMPEXTENT
DATASET IDENTIFICATION


Name of the quality test applied to the data.
LINEAGE, QUALITY & ASSOCIATION
DIC_MET_QUALITY_MEAS
DATA QUALITY
information on quality tests on the dataset, either quantitative measures or qualitative conformance tests.
SERVICE OPERATES ON
Information about a dataset that the service operates on.
DATASET IDENTIFICATION*
* '*lineage_statement*' column in this table holds lineage information.

2.3 EXAMPLE OF SCRIPTS

CreateDB_Oracle_v1_0.sql *

```
-- Terms of Use:  Free for commercial and non-commercial use.  Please acknowledge the material  as:
-- "Copyright/database right, reproduced with the permission of the British Geological Survey © NERC 2011.  All rights reserved".
--
-- ER/Studio Data Architect 9.0 SQL Code Generation
-- Company :      BGS
-- Project :      BoreholeDatabase_ERS9.dm1
-- Author :       BGS
--
-- Date Created : Thursday, September 29, 2011 11:55:50
-- Target DBMS : Oracle 10g
--
-- TABLE: BHD_INDEX
--

CREATE TABLE BHD_INDEX(
    BHD_INDEX_ID              NUMBER(38, 0)    NOT NULL,
    QS                        VARCHAR2(6)      NOT NULL,
    RT                        VARCHAR2(2)      NOT NULL,
    NUMB                      NUMBER(5, 0)     NOT NULL,
    BOREHOLE_SUFFIX           VARCHAR2(4)      NOT NULL,
    BOREHOLE_NAME             VARCHAR2(100)    NOT NULL,
    XY_SOURCE_CODE            VARCHAR2(5),
    CONFIDENTIALITY_CODE      VARCHAR2(5)      DEFAULT '?',
    INCLINATION_TYPE_CODE     VARCHAR2(5)      DEFAULT '?' NOT NULL,
    DRILLED_DATE              DATE,
    DRILLED_DATE_PREC_CODE    VARCHAR2(2)      DEFAULT '?',
    DRILLED_LENGTH            NUMBER(7, 2),
    START_POINT_TYPE_CODE     VARCHAR2(5)      DEFAULT '?' NOT NULL,
    START_HEIGHT              NUMBER(7, 2),
    START_HEIGHT_PREC_CODE    VARCHAR2(5)      DEFAULT '?',
    DRILLING_METHOD_CODE      VARCHAR2(5)      DEFAULT '?' NOT NULL,
    CLIENT_CODE               VARCHAR2(5)      DEFAULT '?' NOT NULL,
    PURPOSE_CODE              VARCHAR2(5)      DEFAULT '?',
    COMMENTS                  VARCHAR2(255),
    X                         NUMBER,
    Y                         NUMBER,
    EPSG_CODE                 VARCHAR2(20),
    XA                        NUMBER,
    YA                        NUMBER,
    PRIMARY KEY (BHD_INDEX_ID)
);


COMMENT ON TABLE BHD_INDEX IS 'Index table of borehole records in the Borehole database'
;
COMMENT ON COLUMN BHD_INDEX.BHD_INDEX_ID IS 'Unique Identifier for a borehole populated by a sequence. This is a primary key of thi
;
COMMENT ON COLUMN BHD_INDEX.QS IS 'Sheet quadrant, ie NE, SE, SW, NW. This is specific to UK borehole record.'
;
COMMENT ON COLUMN BHD_INDEX.RT IS 'This refers to Record Type. For example, B for Borehole Journal etc. This is specific to UK bore
;
```

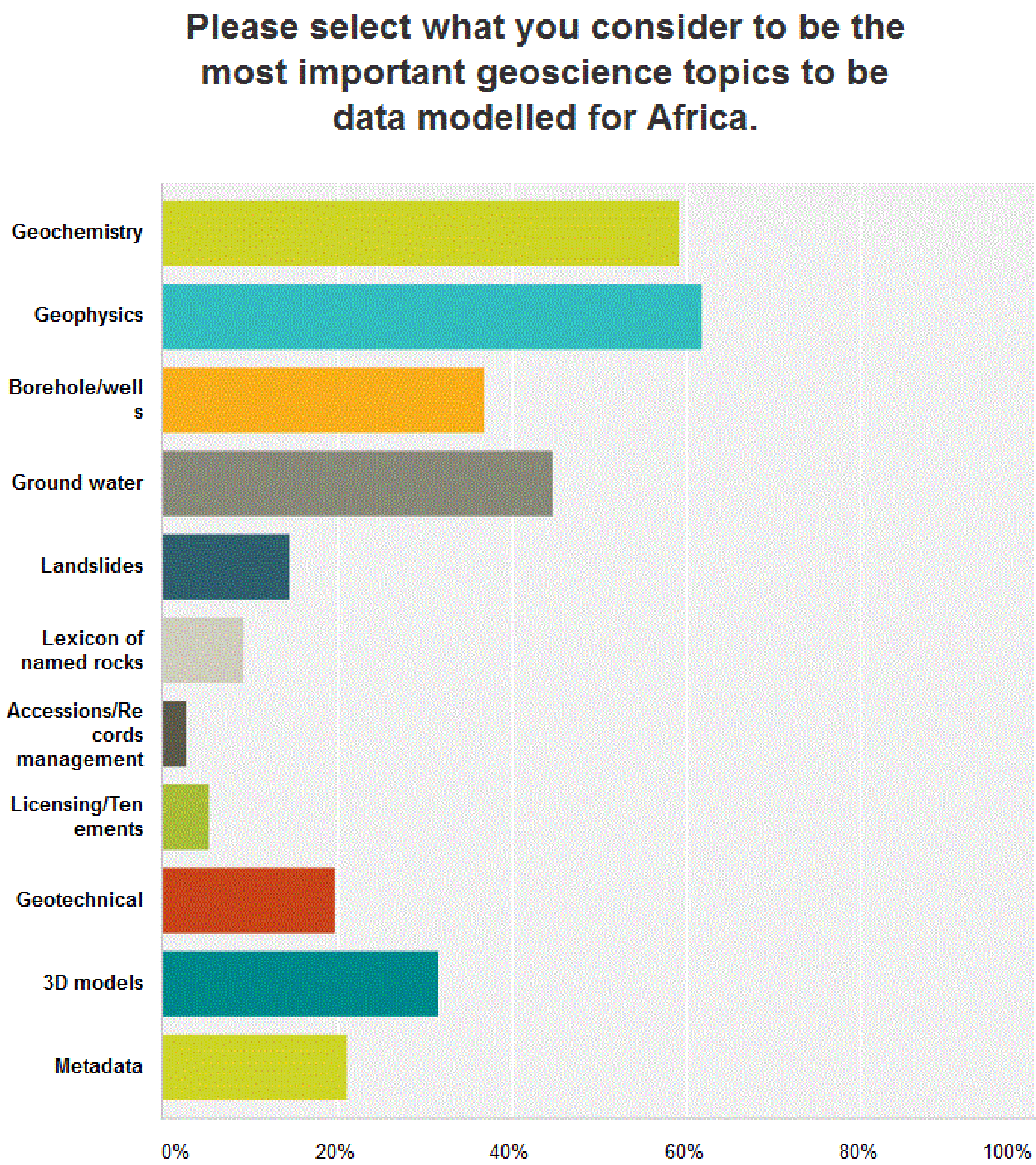
Sample of the SQL scripts for the creation of Borehole data model in an Oracle implementation

InsertScripts_Oracle_v1_0.sql

```
--
-- DIC_BASE_BED --
--
SET DEFINE OFF;
Insert into DIC_BASE_BED
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('!', 'Not Available - Applicable but try as we might,we can''t find a value', 'Not Available', 'C');
Insert into DIC_BASE_BED
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('?', 'Not Entered - We haven''t assigned a value yet', 'Not Entered', 'C');
Insert into DIC_BASE_BED
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('-', 'Not Applicable', 'Not Applicable', 'C');
Insert into DIC_BASE_BED
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('AC', 'Angular Unconformity', 'Angular Unconformity', 'C');
Insert into DIC_BASE_BED
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('ES', 'Erosion Surface', 'Erosion Surface', 'C');
COMMIT;


--
-- DIC_BOREHOLE_DIRECTION --
--
SET DEFINE OFF;
Insert into DIC_BOREHOLE_DIRECTION
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('!', 'Not Available - Applicable but try as we might,we can''t find a value', 'Not Available', 'C');
Insert into DIC_BOREHOLE_DIRECTION
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('?', 'Not Entered - We haven''t assigned a value yet', 'Not Entered', 'C');
Insert into DIC_BOREHOLE_DIRECTION
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('-', 'Not Applicable', 'Not Applicable', 'C');
Insert into DIC_BOREHOLE_DIRECTION
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('S', 'Borehole drilled from ground surface', 'Surface', 'C');
Insert into DIC_BOREHOLE_DIRECTION
   (CODE, DESCRIPTION, TRANSLATION, STATUS)
 Values
   ('U', 'Borehole drilled underground', 'Underground', 'C');
COMMIT;
```

Sample of the SQL scripts for the population of an Oracle implementation of the borehole data model, the data includes example dummy values as well as content for the dictionaries of controlled terms used in the design.



Example of the results from the questionnaire handed out at the 24th meeting of the Colloquium of African Geology (CAG24)

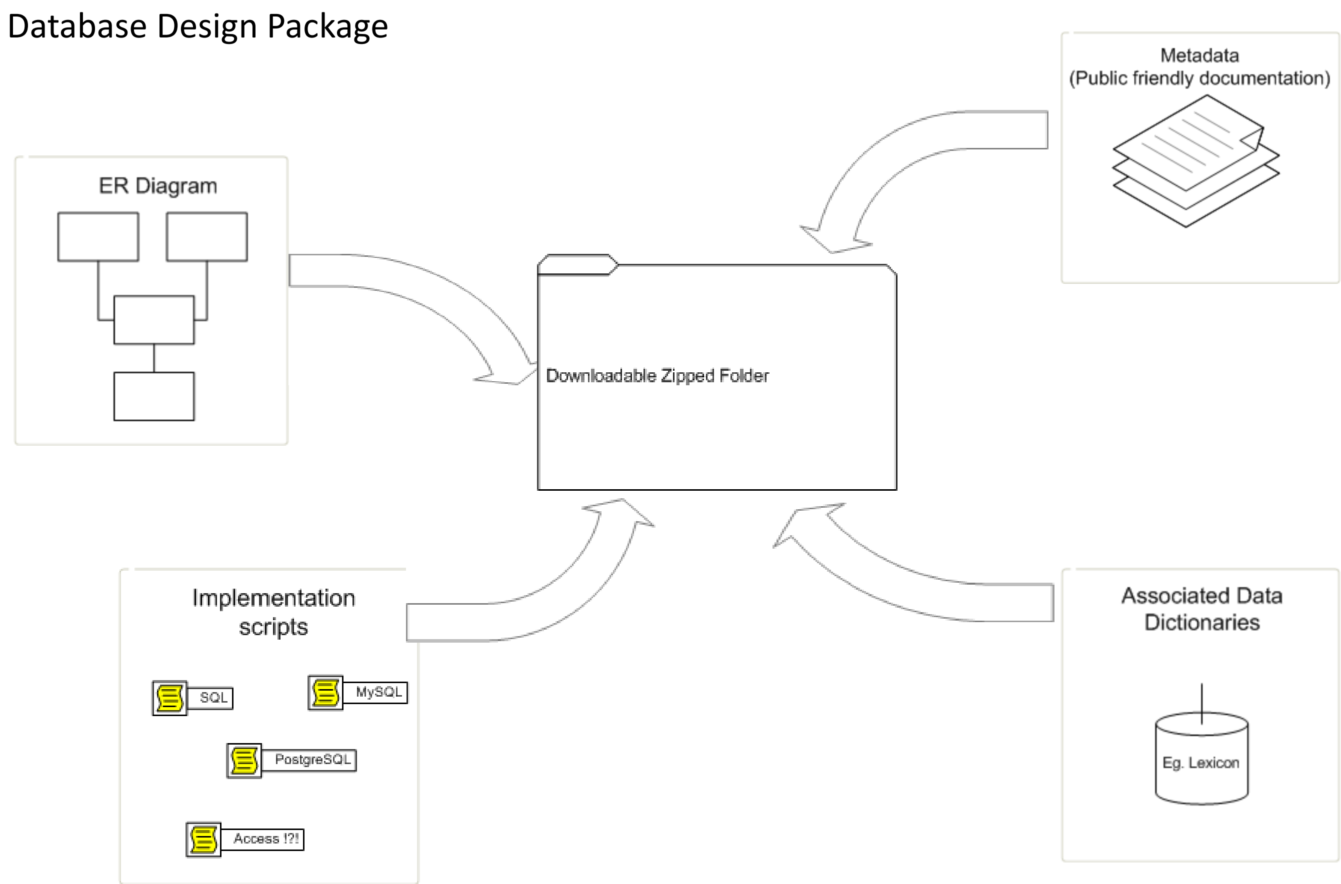


Diagram to show the components which make up a downloadable database design package