Lexicon of Named Rock Units Data Model Summary

For the Open Geoscience Data Models Project

Baneet Arora & Carl Watson

BGS Database Solutions Team

Contributions from:

Martin Nayembil, Tim McCormick

1. Introduction

We believe that most readers will imagine the data model described in this document as a physical database, therefore we use the terms 'Table' and 'Column' where 'Entity' and 'Attribute' would be more technically correct.

1.1. Document purpose

This document provides a summary of the Lexicon of Named Rock units Data Model produced by the BGS for the Open Geoscience Data Models project. It includes descriptions of the main components (also known as tables or entities) which make up the data model, the relationships between these components and the dictionaries which are, in effect, controlled vocabularies of the terms used in describing, and supporting the description of, scientific and other observations.

This document does not deal with technical details relating to specific database hardware or software, for those please see the implementation specific download.

1.2. Simplified data model

The BGS Lexicon of Named Rock Units Database holds information about all allostratigraphical, lithostratigraphical and lithodemic units of member (or equivalent) and higher rank that are used, or have been used, on BGS maps and in BGS publications. It includes information about lesser named units, notably beds (and their equivalents), and some information about names not currently recognised by the BGS or now regarded as obsolete. In this download we have provided a simplified data model based upon the current BGS Lexicon of Named Rock Units Database.

We have simplified the model to some extent because we want to highlight the essence of the database but not to present an overly complex design. We have removed some features and tables which relate to concepts which are specific to the BGS and which may not be relevant to other organisations.

We hope that this design can provide one example of how you might structure a Lexicon data model while highlighting areas that require further consideration when designing a lexicon database.

1.3. BGS standards and naming conventions

The creation of the model has been carried out by BGS staff using corporate database standards and best practices.

The table and column names have been altered to more generic versions of those in use on the BGS corporate database, however, the names used in this data model conform to the BGS database naming conventions, for example all dictionary names should begin with DIC_.

1.4. A brief history of the BGS Lexicon of Named Rock Units Database

The following bullet points are based upon BGS Reports and discussions with current BGS staff:

- Prior to the existence of the database version of the Lexicon, BGS used a 'Dictionary of Stratigraphy' that was built to hold unique codes for all rock units shown on BGS maps and referred to in publications. This dictionary table held a much restricted set of information and was backed up by paper records.
- The first database hosted version of the Lexicon was developed in 1990 and it allowed BGS geoscientists to a store wealth of information that defined a rock unit digitally for the first time. The design of the database has been modified from time to time in response to the requirements of BGS geologists.
- The current lexicon database contains more than 19000 unit names, storing definitions and supplementary information on lithostratigraphic and lithodemic units of Member status, or equivalent, and above which are used on BGS maps and in BGS publications, products and services.
- In 2000, the Lexicon was made available on the BGS website through a simple search page; this was followed by a web service which made the lexicon data freely accessible to public in 2008.
- The database was re-engineered and a new version released in 2012. The new version enhanced the existing design and this was done by adding new or altering existing tables to enable greater linkages with other databases within BGS. The new design allows geologists to capture lots more information which was not possible earlier.
- The database is dynamic as new entries can be added, existing entries can be revised, upgraded and reclassified using BGS internal Lexicon Data Entry System.

For more information on the history of the BGS Lexicon of Named Rock Units Database see the publication:

McCormick, T. 2011. The British Geological Survey Lexicon of Named Rock Units. In: Foote, J B (ed.) Navigating the Geoscience Information Landscape: Pathways to Success. *Proceedings of the Geoscience Information Society*, Vol. 40 (for 2009), 47-53.

1.5. The BGS Lexicon workflow

Before we look at the details of the data model this section will briefly describe a typical workflow that would result in data being entered into a lexicon database within BGS.

- The application for a new entry to be created in the Lexicon is done by emailing either the Lexicon Manager or the Information Management Project Leader.

- Enough information is provided to the Lexicon Manager to help him determine if the unit already exists in the Lexicon (possibly under another name), and to enable the Lexicon Manager to create a "stub" entry (with basic details) in the Lexicon. The information provided must include the following:

  - Rock unit name
  - Rock unit rank
  - Onshore/Offshore/Both
  - Name of the parent unit
  - Any alternative/previous names by which the unit is/has been known.
  - Age (you can specify a range between a minimum and maximum extent , if necessary)

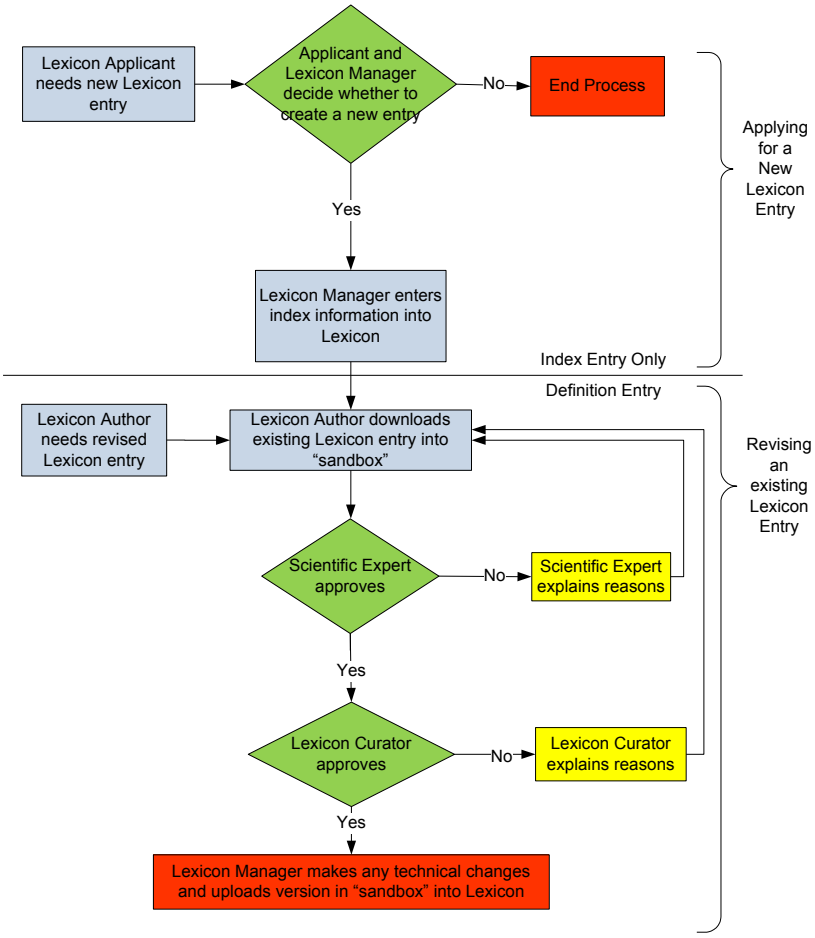

Figure 1

In addition,

- An abbreviation code for the unit can be suggested
- Lithology can be described

Based on the information provided, the Lexicon Manager will determine if the new entry is required and if so, will create a stub entry in the Lexicon of Named Rock Units.

- If the new entry is at Member rank, equivalent or above, it will be created as a "Pending Upgrade" entry, meaning that the code and minimal information are present but needs enhancement. It is the applicant's responsibility to turn this into a full Lexicon entry using the data entry system with other details.

- If the new entry is below Member rank or equivalent, it will be created as a “Index Level” entry, meaning that no further information is required. Figure 1 shows the workflow for applying for a new Lexicon entry and revising already existing entry.

2. The data models explained

2.1. Current BGS Lexicon of Named Rock Units Database

The current BGS Lexicon Database has evolved over many years, it will not be described in detail in this document, however you can get a feel for the latest version of BGS Lexicon data model from the diagram below.

Figure 2

- The main tables (recording lexicon unit definitions, parent unit, equivalent names, lithologies, type localities, references and other information) are shown in yellow.
- The tables that are part of the approval process of the lexicon database are shown in white. They hold approval information for a definition in the Lexicon of Named Rock Units, the current status of a step in the approval process for a new or revised entry, information about parties involved in creating and approving definitions in the Lexicon.
- The dictionaries that are specific to lexicon database are shown in grey.
- Other dictionaries which are not specific to lexicon database are shown in green.

For more information about the BGS Lexicon Database and its uses, there are a range of publications available online or via the NERC Library. Some are listed in the section - A brief history of the BGS Lexicon database.

2.2. The Lexicon of Named Rock Units Data Model for the Open Geoscience Data Models project

For more technical details and a more complete "Entity-Relationship" Diagram see the technical documentation that accompanies the implementation code.

The following diagram shows the new and simplified logical design which has been created as part of the Open Geoscience Data Models project:

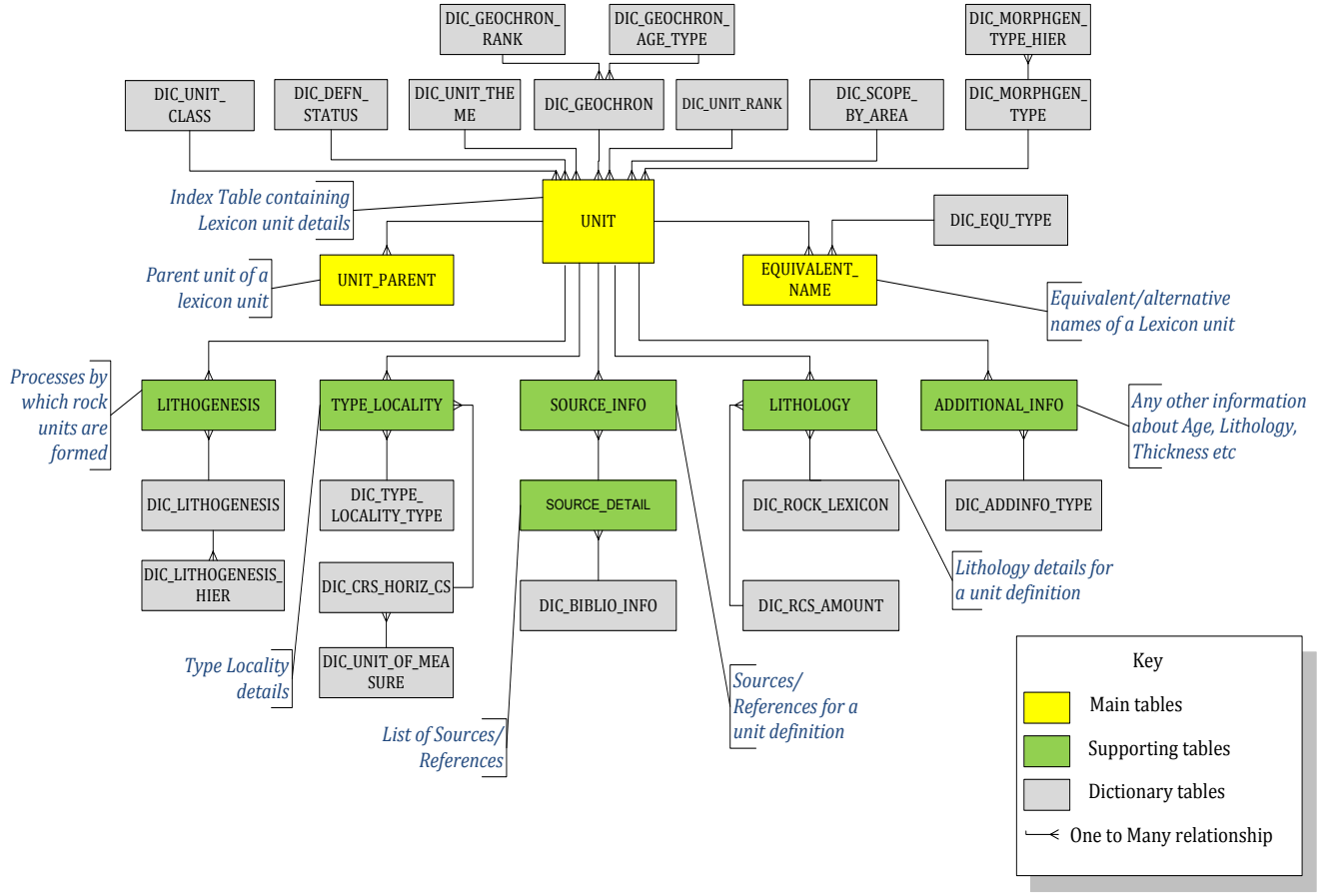

Figure 2

2.3. Components of the Data Model

The data model is composed of three types of tables, they are:

i. Main tables containing index level and non-index level information about a lexicon unit.
ii. Supporting tables that record information about Type Locality, Lithology, References, Lithogenetic processes and any other additional information about a lexicon unit.

iii. Dictionaries (controlled vocabularies) used to constrain the data, i.e. restrict the data that may be entered for certain values.

Naming Conventions

It is standard practice in BGS corporate database design to assign a three letter prefix to table names to indicate the general subject areas to which these tables belong. In this case we have used the prefix 'LXN_' to represent the lexicon database.

Lexicon Unit Index Table: LXN_UNIT

This is the main table of the Lexicon of Named Rock Units database, containing index level and non-index information about each geological unit defined in Lexicon. This table should be populated with the mandatory rock unit details before other attributes can be captured in other tables.

Typical information to be captured in this table includes:

- Rock unit name
- Unit Rank and Unit Class
- Unit Theme
- Definition Status
- Whether the unit definition is Formal or Informal
- Whether the unit name is used in onshore or offshore context or both
- Details of the chronostratigraphy of the unit, etc.

Lexicon Unit Parent Table: LXN_UNIT_PARENT Naming Conventions

This table holds the parent-child relationships between the geological units defined in the Lexicon of Named Rock Units.

Most geological units have at least one "Parent Unit". There are some that have more than one parent. Thus, the relationship between LXN_UNIT and LXN_UNIT_PARENT is one-to-many. For example, *Aegiranum Marine Band* lexicon unit has three parent units i.e. *Pennine Middle Coal Measures Formation*, *South Wales Middle Coal Measures Formation* and *Scottish Upper Coal Measures Formation*.

Lexicon Equivalent Name Table: LXN_EQUIVALENT_NAME

This is a linking table that is used to record the equivalent or alternative names for a geological unit as part of the Lexicon of Named Rock Units.

An "Equivalent name" can result because a name becomes obsolete, because of partial overlap, because there are lateral equivalents (partial or total), or it may just be a different name for the exact same geological unit. Each Lexicon definition may include none, one, or many equivalent names, making the relationship between LXN_UNIT and LXN_EQUIVALENT_NAME one-to-many. For example, *Aegiranum Marine Band* lexicon unit has many alternative and obsolete names like *Bolton Marine Band*, *Chance Pennystone Marine Band*, *Cefn Coed Marine Band*, *Crofts End Marine Band* etc.

Source Detail Table: LXN_SOURCE_DETAIL

This table holds a list of sources of information, usually articles in the scientific literature, used to create definitions in the Lexicon of Named Rock Units. Reference details sourced from other data stores can also be recorded in this table, linking to the original data store by an identifier. For example, an ID Number from an OLIB Library Management System or an Endnote file can be recorded in this table.

Information to be captured in this table includes:

- Identifier of the External Store
- The literature reference
- Data store from which the record was sourced. For example, OLIB.

Lexicon Source Table: LXN_SOURCE_INFO

This table links Lexicon unit definitions to the literature citations or other information sources. There may be many citations from LXN_SOURCE_DETAIL for each lexicon definition. Thus, this table resolves the many-to-many relationship between LXN_UNIT and LXN_SOURCE_DETAIL into two 'one-to-many' relationships, with each of the tables acting as the 'one'.

Lexicon Type Locality Table: LXN_TYPE_LOCALITY

In this table, you record information on the type localities, type areas and other significant reference localities for the geological units in the Lexicon of Named Rock Units. Each Lexicon definition may include one or more type localities. Some classes of geological unit have no type locality, for example mass flow and artificial deposits. The relationship between LXN_UNIT and LXN_TYPE_LOCALITY is one-to-many.

Typical information includes:

- Kind of Type Locality. For example, Type Section, Type Area, etc.
- Minimum and Maximum X and Y coordinates
- Accuracy of X and Y coordinates
- Description of the type Locality

Lexicon Lithology Table: LXN_LITHOLOGY

This table holds coded Lithology values used in Lexicon of Named Rock Units. Each Lexicon definition should have one or more "Lithology" values, thus making the relationship one-to-many.

This table contains a column to record rock code that represents the Lithology which is constrained against a dictionary table. It is also possible to record the amount of Lithology that is present. For example, values like Main, Subsidiary, and Trace can be used. This column is constrained against a dictionary table. In addition, there is a column to flag whether the Lithology is locally dominant in any part of the geological unit.

Lexicon Lithogenesis Table: LXN_LITHOGENESIS

In this table, you record processes and environments by which rocks and other geological deposits are formed, for the Lexicon of Named Rock Units. Each Lexicon definition may include one or more process. Hence, the relationship between LXN_UNIT and LXN_LITHOGENESIS is one-to-many.

This table contains a column, which is constrained against a dictionary table, to record a code representing a lithogenetic process.

Lexicon Additional Information Table: LXN_ADDITIONAL_INFO

This table can be used to hold additional/free text information for the definitions in the Lexicon of Named Rock Units. There can be several types of free text associated with a particular definition. Therefore, the relationship between LXN_UNIT and LXN_ADDITIONAL_INFO is one-to-many.

There is a dictionary constrained column in this table that stores the 'type' of free text information. For example, Lithology, thickness, definition of upper boundary, definition of lower boundary, commonly associated landforms, etc. The table has another column that holds the actual free text information.

3. Points of interest and suggested changes

This section contains a few points we wish to highlight as we believe they could help anyone wishing to use the data model presented earlier in a real world situation.

The data model we have presented is only a basic version and requires alterations, additional columns and tables to suit local or organisation specific requirements, the following comments are intended to stimulate thoughts on what those changes could involve.

Separate index level information from non-index information

In the BGS Lexicon of Named Rock Units, we have separated the index level information, from the non-index information into two tables. Index Level information is the mandatory information required to define a lexicon unit. For example, Unit code, Unit name, Unit ranks, Unit class. Non-Index level information is the supplementary information that defines a unit. For example, Age, Litho-morpho-genetic type, etc.

For the sake of simplicity, the data model presented in section 2.2 combines the index and non-index level information in one table i.e. LXN_UNIT.

However, if you want to record the index level information and non-index level separately, then you may consider adding another table (we have called this table LXN_UNIT_FULL in the BGS Lexicon of Named Rock Units Database) and add the columns that represent non-index level information to this table.

If implemented as above, there needs to be an index level entry in LXN_UNIT table before any non-index data can be recorded in LXN_UNIT_FULL table. There can only be zero or one record in LXN_UNIT_FULL table for every record in LXN_UNIT table, therefore the relationship between the two tables is one-to-zero or one. The diagram below shows the relationship between the two tables and the columns that will become part of the LXN_UNIT_FULL table should you want to separate the non-index level data from the index level data.

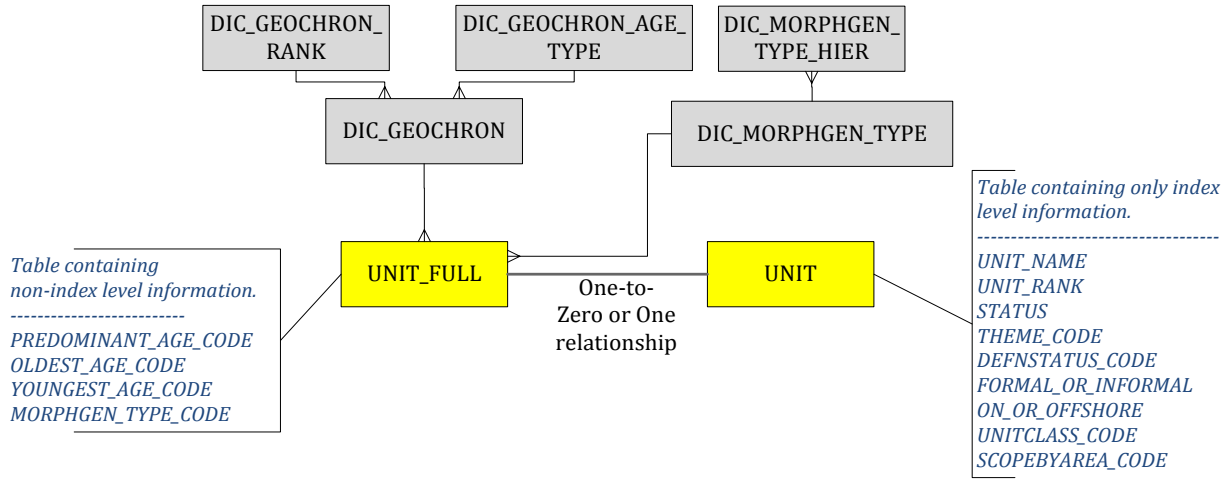

Figure 3

Tables to track revisions/approvals for lexicon unit definitions

The data model we have provided can be extended to track the revisions/approvals for lexicon unit definitions. In the BGS Lexicon Database we have set of tables that hold approval information and the following diagram and table descriptions show an example of how the data model can be updated to capture these details.

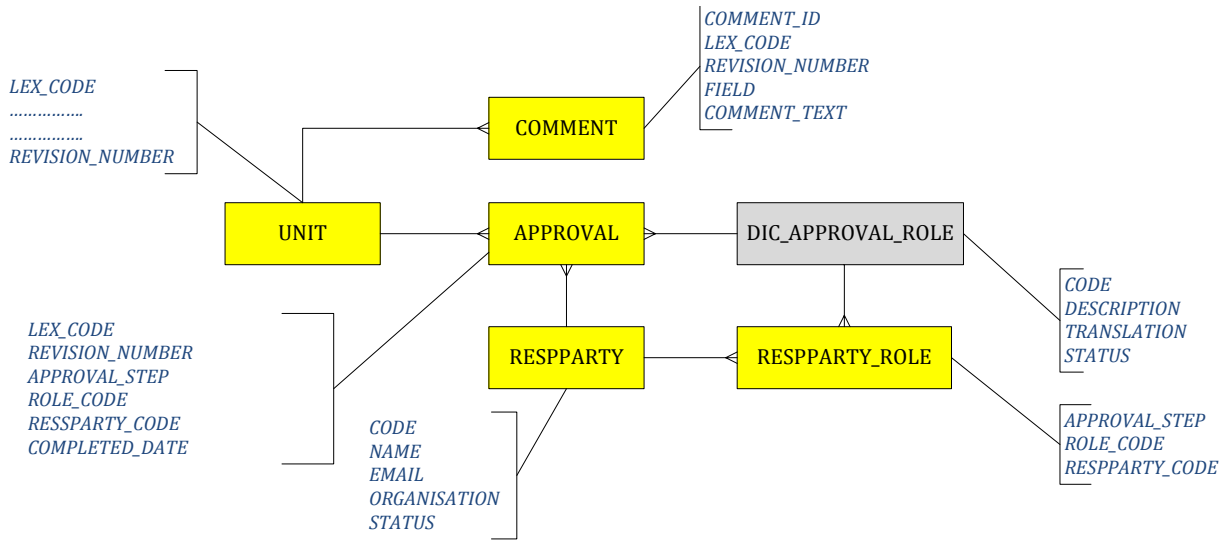

Figure 4

DIC_APPROVAL_ROLE

This will be a controlled list of roles in the approval process for creating and editing definitions in the Lexicon of Named Rock Units. In BGS, the roles are Lexicon Applicant, Lexicon Author, Lexicon Curator, Scientific Expert and Lexicon Manager. However, this

dictionary can be updated with values that specify organisations individual approval chains.

LXN_APPROVAL

This table will hold approval information for a definition in the Lexicon of Named Rock Units. There will be several rows in the table for each revision of an entry. Each row will indicate the current status of a step in the approval process for a new or revised entry. For example, in the approval chain: Step 0 = Lexicon Applicant; Step 1 = Lexicon Author; Step 2 = Lexicon Curator; Step 3 = Scientific Expert; Step 4 = Lexicon Manager. Referring to Figure 4,

- REVISION_NUMBER column will contain the version number of the definition. The first definition of a unit is version 0. The next version is 1, etc.
- APPROVAL_STEP column will represent the step in the approval chain. Steps can be numbered in sequence within the version number. Application for a new lexicon entry by a Lexicon Applicant is step 0. Only version number 0 of a unit will have a step 0 (because you do not apply for a new code when revising an existing definition).
- ROLE_CODE column will identify the role (lexicon applicant, lexicon author, scientific expert, etc.) carried out by the responsible party at this step of the approval chain. This will be constrained against dictionary table DIC_ROLE.
- RESSPARTY_CODE column will identify the party responsible for this step of the approval process. This column will be constrained against LXN_RESPPARTY table.
- COMPLETED_DATE column will hold the date on which this step of the approval process was completed.

LXN_RESPPARTY

This table holds list of responsible parties involved in creating and approving definitions in the Lexicon of Named Rock Units. The table will include columns to store name, email, organisation, status of the responsible party etc. This table can be modified according to individual organisational needs.

LXN_RESPPARTY_ROLE

This table lists who the allowable parties are for each step and role in the approval process for creating and revising definitions in the Lexicon of Named Rock Units. This table does not control who may apply for a code or write/revise a definition, because these steps may be carried out by any geoscientist. For example, this table will have a row with APPROVAL_STEP as 4, ROLE_CODE as LEXMANAGER and RESPPARTY_CODE as the code from LXN_RESPPARTY that identifies the actual lexicon manager.

LXN_COMMENT

This table will hold comments on the different attributes for various revisions of a lexicon unit definition. This table has a column REVISION_NUMBER that will hold a revision no. of a lexicon code. The FIELD column in figure 4 will be used to store a code to identify the attribute. The attributes can be equivalent name, reference, age, Lithology, approval, locality, Lithogenesis etc. The COMMENT_TEXT column will hold the actual comment.

If you do decide to extend the data model to track approval/revision information using the above mentioned design, then a column REVISION_NUMBER should be added to LXN_UNIT table to identify the current revision number for a particular unit definition.

Replace non-lexicon dictionary tables with other existing tables

In this data model, some of the tables are linked to dictionaries which are BGS standard controlled vocabularies or are part of other databases. These are not lexicon specific dictionaries. You don't necessarily have to use the same structure for these dictionaries as we have provided in this data model. Below is the list of dictionaries that can be replaced with tables or controlled vocabularies that already exist and are specific to your organisation.

- Dictionary tables holding codes, names and radiometric ages of Geochronological intervals, their age type and rank that can be used in BGS databases. These tables are DIC_GEOCHRON, DIC_GEOCHRON_AGE_TYPE, and DIC_GEOCHRON_RANK.
- DIC_CRS_HORIZ_CS is a dictionary of horizontal coordinate systems, mostly sourced from EPSG database. Data from the normalised EPSG database is summarised and joined when populating this table.
- DIC_UNIT_OF_MEASURE is a dictionary of measurement units used for a specific property; e.g. metres, centimetres, inches, chains, fathoms, kilograms, tons, pints, quarts, gallons, litres, square metres, cubic metres.
- DIC_ROCK_LEXICON is a dictionary subset of a master dictionary from BGS Rock Classification Scheme Database.
- DIC_RCS_AMOUNT is a dictionary table which holds codes for the amount of single lithologies that make up a composite lithology. This table is also part of BGS Rock Classification Scheme Database.
- DIC_BIBLIO_INFO is a dictionary table of the controlled vocabulary for the source of bibliographic references. E.g. OLIB.

Linkages to other databases

Lexicon of Named Rock Units is one of the key databases in BGS. It has linkages with many other databases. One of the databases that use the Lexicon of Named Rock Units is the Boreholes database. The borehole data model is available through the Open Data Models Project. In the borehole data model, the LITHOSTRAT_CODE column in the BHD_LOG table references the LXN_UNIT table. This LXN_UNIT table records only the index level information about a unit definition. This is how we have implemented it in the BGS (see Figure 3).

But, for the sake of simplicity in implementation of the Lexicon database, we are proposing that you combine the index and non-index information in the LXN_UNIT table. We would leave this to you to decide which approach to choose and how it would fit around your organisational needs.

If you have already implemented the borehole database that we have provided through the Open Data Model Project, then its linkages with the Lexicon of Named Rock Units is fairly easy as demonstrated in the diagram below

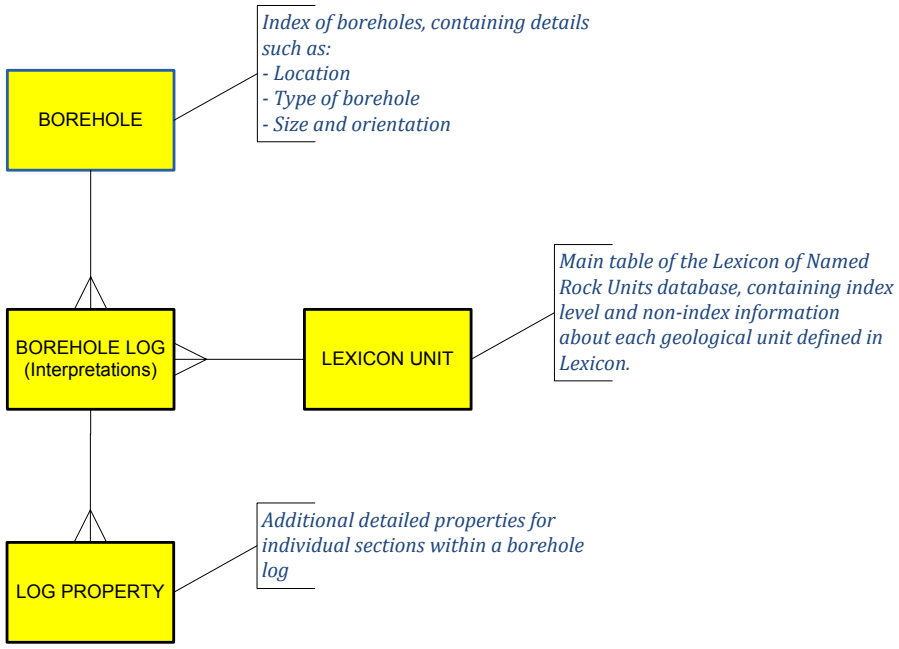

Figure 5

History tables and auditing procedures

In BGS, we have a standard way of auditing the data in our databases. We record details of person entering, updating or deleting a record. All the tables have four audit columns i.e. USER_ENTERED, DATE_ENTERED, USER_UPDATED, DATE_UPDATED.

When a new row is inserted in the master table, the USER_ENTERED and DATE_ENTERED columns are populated with the oracle user id of the person doing the insert and the current date and time respectively captured by a trigger placed on the table.

We hold previous copies of data through History tables, sometimes referred to as Audit tables. They serve the purpose of holding the old versions of table rows that have subsequently been modified or deleted. A History table is associated with one, and only one master table. It is named by appending the string _HIST to the name of the master table.

These history tables are populated by a trigger placed on the master table. When an existing row in the master table is about to be modified, a copy of that row is inserted into the history table before any modifications are applied to it. As the row is moved to the History table, three additional audit columns i.e. THEFUNC, THEUSER, THE DATE are populated. The THEFUNC is populated with the letter 'U' for update, 'D' for delete; THEUSER is populated with the oracle user id of the person applying the update, and THEDATE is populated with the current date and time. Once the above is completed, the master table row is updated; USER_UPDATED and DATE_UPDATED columns on the master table are populated with the user id and current date and time respectively.

By using the History table system, the previous contents of a Master table row can be examined, for as far back in time as the auditing process has been in operation and preserved. Furthermore, any accidental deletions or updates can be corrected, if required.

The above gives an overview of how auditing is done within BGS, you may want to follow similar procedures to audit your data or use a corporate standard that exist within your organisation.