Borehole Data Model Summary

1. Introduction

1.1 Document purpose

This document provides a summary of the Borehole Data Model, it includes descriptions of the entities, relationships as well as the dictionaries (controlled glossary of terms) used to constrain the data.

This document does not deal with database platform specific technical details, for those details please see the additional documents in the implementation folder.

The data model contains entities and attributes that you may consider optional or in need of some alteration to most suit the needs of your organisation. The most obvious of these have been highlighted, for example in Appendix 3 table and column names appear in **bold** text unless they are considered to be either optional or likely to require some alteration to best meet the needs of organisations outside of the BGS. In addition, section 3 also describes some of the possible enhancements you may wish to make to the design.

1.2 Borehole data model

This borehole data model is based upon the BGS database for index level data on the existence and location of onshore boreholes.

"The Single Onshore Boreholes Index (SOBI) is an index of over one million records of boreholes, shafts and wells from all forms of drilling and site investigation work held by the BGS.

The collection covers onshore and near shore boreholes from Great Britain dating back to at least 1790 and ranging from one to several thousand metres deep."

The index level table provided in this data model is capable of storing information regarding the position, who drilled the borehole, its purpose and confidentiality rating. The data model also contains additional entities that enable the capture of multiple geological interpretations of each borehole log.

1.3 BGS standards and naming conventions

The creation of this borehole data model and associated documentation has been carried out by BGS staff using corporate standards and best practices.

The entities and attribute 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 dictionaries should begin DIC_. For more information about BGS database naming conventions please see Appendix 2.

2. Data Model Description

2.1 High level cartoon

This following diagram is a high level cartoon of the borehole data model, for a full version of the model entity-relationship diagram see Appendix 1.

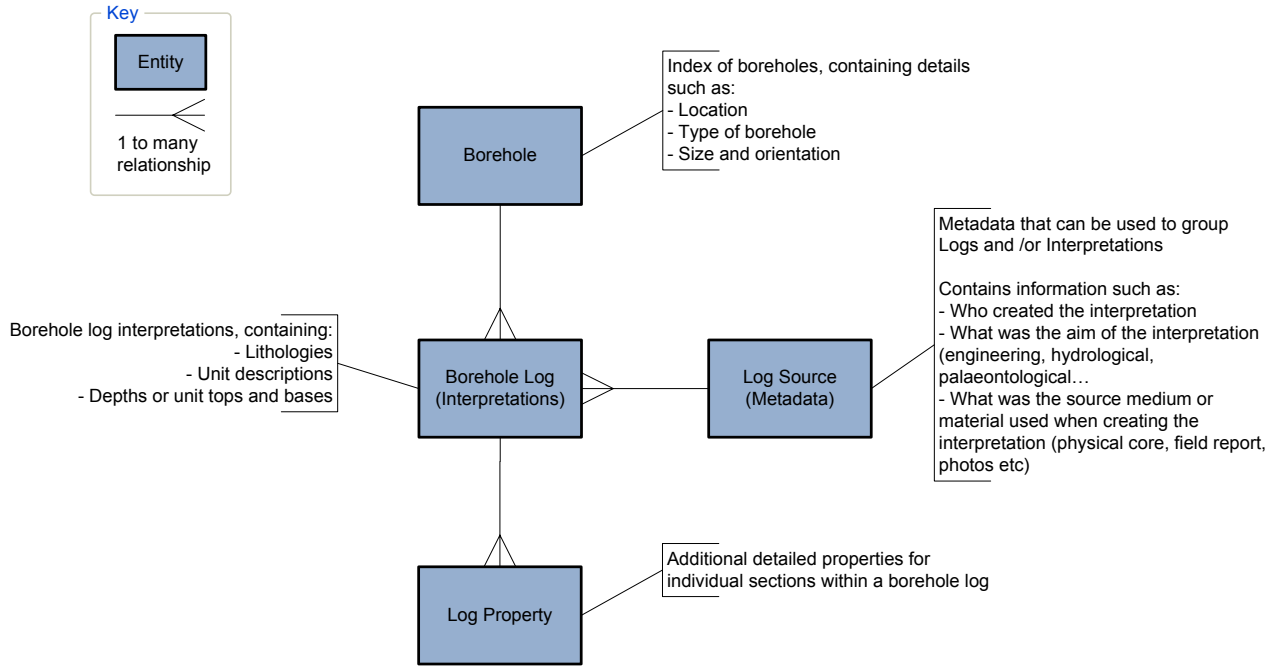


Fig. 1

2.2 Data Model Entities

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

i. Main tables, which are essential and contain the core details about a given borehole and interpretations of it.
ii. Other tables, provide useful metadata on the borehole records captured in the main tables.
iii. And finally, the dictionaries (controlled vocabularies) which are used to constrain the data.

Main tables

- BHD_INDEX (Shown in Fig. 1 as 'Borehole')
  An index table of borehole records within the Borehole database.
- BHD_LOG (Borehole interpretations log)
  A table within the Borehole database that holds geological information of the borehole log.
- BHD_LOG_PROPERTY (Log property)
  A table within the Borehole database that holds the borehole log property data.
- BHD_LOG_SOURCE (Interpretation meta data)
  Grouping table for borehole logs (includes interpreter, project, source medium)

Other tables

- LXN_UNIT
  The main table used within the BGS lexicon database containing index level information about the Lexicon Unit

- STAFF
  Table to store a list of staff and their details.

  ** The attributes used in this dictionary are likely to vary from one organisation to another, and it may be desirable to replace this table with a link to an existing staff table within your organisation.
- PROJECT
  Table to store list of projects and their details.

  ** Within the BGS most work tasks are associated with a 'Project' and these are all listed in a single table. When users of the boreholes database are selecting which interpretations they want to see for a given borehole they might filter them according to the project which was responsible for their creation.

  The attributes used in this dictionary are likely to vary from one organisation to another and you may not need an equivalent table at all.
- DIC_SOURCE_MEDIUM
  Dictionary of codes to describe source mediums/materials
  The list of source materials that borehole interpretations could be based on, for example physical core, field reports or photos and as with the Project table this can be used when selecting which interpretations are of most value for a given borehole.

Dictionary/controlled vocabularies

- DIC_BASE_BED
- DIC_BOREHOLE_DIRECTION
- DIC_COMPANY
- DIC_CONFIDENTIALITY
- DIC_DATE_ACCURACY
- DIC_DRILLING_METHOD
- DIC_DRILLING_PURPOSE
- DIC_INCLINATION
- DIC_ORD_DATUM_PREC
- DIC_PROPERTY_TYPE
- DIC_RELIABILITY
- DIC_ROCK_ALL
- DIC_UNIT_OF_MEASURE
- DIC_XYSOURCE

2.3 Key Relationships

- A Borehole can have zero or more log record but a particular individual log is about one borehole only. Therefore, the relationship between borehole (BHD_INDEX) and borehole log (BHD_LOG) is one to zero or many.

- A borehole log can define zero or more properties and a particular property type can exist in zero or more logs. Therefore, the relationship between borehole log (BHD_LOG) and property (DIC_PROPERTY_TYPE) is many to many. Therefore, this relationship is broken into one to many and many to one by adding a link table that stores borehole log property data i.e BHD_LOG_PROPERTY.

- A particular log source (set of interpreter, project and source) can record zero or more logs and one individual log can only have one log source. Therefore the relationship between log source (BHD_LOG_SOURCE) and log (BHD_LOG) is one to zero or more.

3. Possible Enhancements

This section contains a series of suggested enhancements that could be made to the borehole data model as earlier in this document.

- **Table to store alternative names for Borehole**: The database can be enhanced by adding a new table that can be used to store alternative/other names of the borehole. The relationship between borehole and alternative name is one to zero or many i.e. each borehole can have none or many alternative names, but each alternative name can be of a single borehole only. Suggested name : **BHD_ALTERNATIVE_NAME**

- **Table to store additional dates for borehole:** A new table to record additional dates can be added to this design. The relationship between borehole and dates is one to zero or many i.e. each borehole can have none or many date records but each date record is for a single borehole. Suggested name : **BHD_EXTRA_DATE**

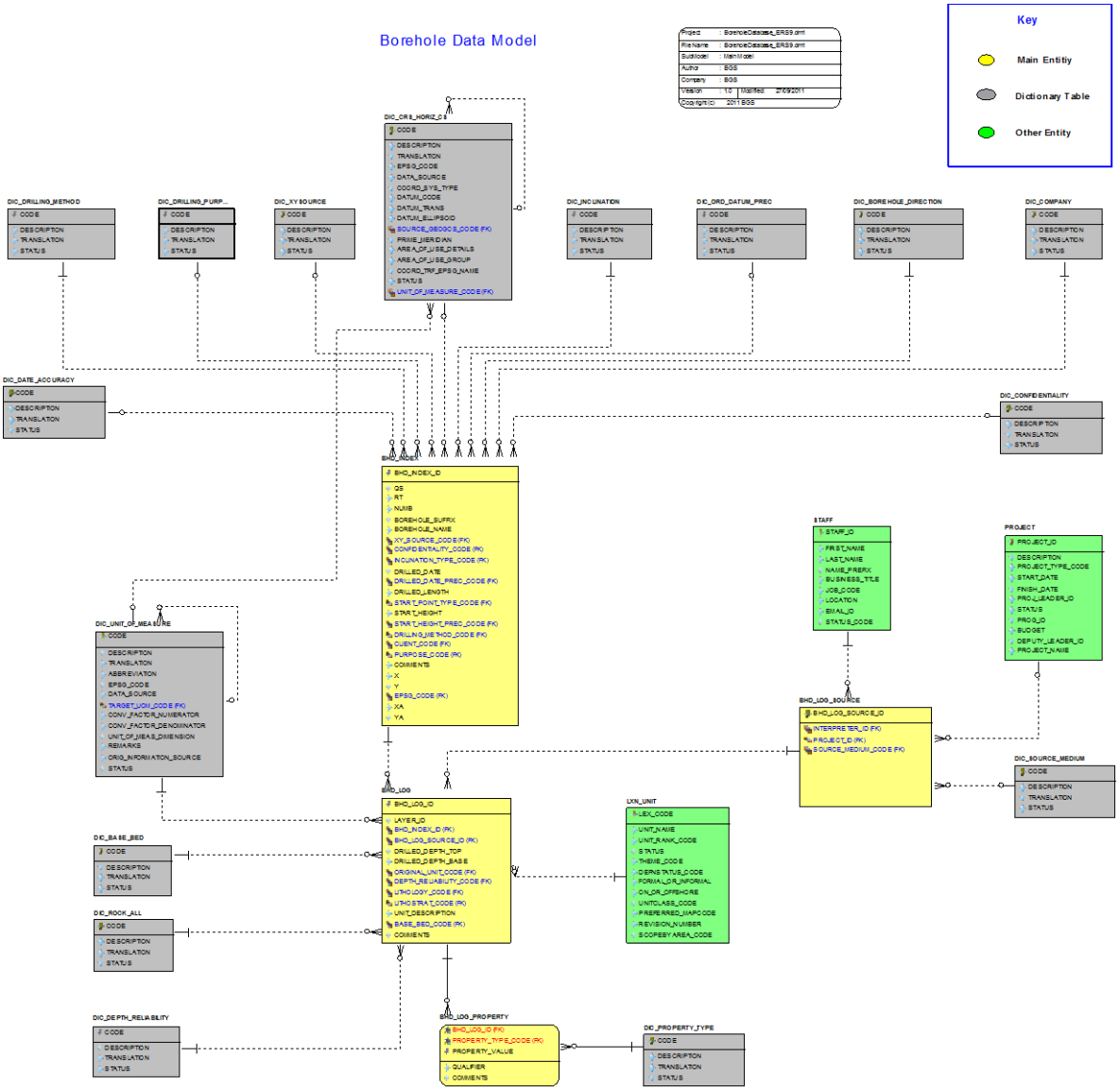
- **Validation Flag fields can be added to any table:** Validation fields can be added to the tables that would allow to record if a particular property/measurement has been validated or not. For example, **VALIDATION_FLAG** field can be added to **BH_LOG** table to indicate if the borehole geology has been validated.

- **History Tables and Auditing procedures:** The History tables, sometimes referred to as Audit tables, serve the purpose of holding 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. Within the BGS Oracle database history tables are populated by a trigger placed on the Master.

Add more, see: http://bgsintranet/scripts/oracle/devGuDoc/guidelinesOverview.cfm

Appendix 1

Entity Relationship diagram:



Appendix 2

Naming Conventions

Dictionary Table	The object name is prefixed by 'DIC_' to indicate that it is a dictionary table. For e.g. DIC_CONFIDENTIALITY
Core table	All the core tables in Borehole Database are prefixed by 'BHD_'. A maximum of 3 characters followed by '_' is used to indicate the main area to which a table belongs. For e.g. BHD_INDEX
History Table	History (audit) table is named after the base table suffixed by 'HIST_' For e.g. BHD_INDEX_HIST
Primary Key	Primary key is named after the table suffixed by '_PK'. For e.g. Primary key on BHD_INDEX table is BHD_INDEX_PK
Foreign Key	Foreign key is named after the table on which foreign key is placed (not the table it references), suffixed by '_FKn' where n is an ascending integer (starting from 1). For e.g. Foreign key on BHD_INDEX is BHD_INDEX_FK1
Unique Constraint (keys)	Unique keys are named after the table, suffixed by '_UKn' where n is an ascending integer (starting from 1). For e.g. Unique key on BHD_LOG is BHD_LOG_UK1
Check Constraints	Check constraints is named after the table, suffixed by '_CKn' where n is an ascending integer (starting from 1). For e.g. Check constraint on DIC_CONFIDENTIALITY table is DIC_CONFIDENTIALITY_CK1
Column that references dictionary	Column that references a dictionary is suffixed by '_CODE'. For e.g. CONFIDENTIALITY_CODE in BHD_INDEX references DIC_CONFIDENTIALITY.

Appendix 3

The following tables contain the table and column descriptions, by default the names appear in **bold** text unless they are considered to be either optional or likely to require some alteration to best meet the needs of your organisation.

BHD_INDEX: An index table of borehole records within the Borehole database.

Column Name	Description
BHD_INDEX_ID	Unique Identifier for a borehole populated by a sequence. This is a primary key of this table.
QS	Sheet Quadrant i.e NE, SE, SW, and NW. This is specific to UK borehole record.
RT	Record type. This is specific to UK borehole record.
NUMB	Running borehole number (limited to a quadrant of a sheet). This is specific to UK borehole record.
BOREHOLE_SUFFIX	Borehole suffix. This is specific to UK borehole record.
BOREHOLE_NAME	Name of the borehole
XY_SOURCE_CODE	Source of geographic coordinate. This is a foreign key to DIC_XYSOURCE.
CONFIDENTIALITY_CODE	Confidentiality status of borehole record. This is a foreign key to DIC_CONFIDENTIALITY.
INCLINATION_TYPE_CODE	Inclination of the borehole. This is a foreign key to DIC_INCLINATION.
DRILLED_LENGTH	Total length of the borehole
START_POINT_TYPE_CODE	Start point of the borehole. This is a foreign key to DIC_BHDR.
START_HEIGHT	Ordnance datum level (surface level of the borehole)
START_HEIGHT_PREC_CODE	Ordnance datum level accuracy. This is a foreign key to DIC_ORD_DATUM_PREC
DRILLED_DATE	Drilling date of the borehole
DRILLED_DATE_PREC_CODE	Accuracy of the drilled date. This is foreign key to DIC_DATE_ACCURACY.
DRILLING_METHOD_CODE	Method used to drill the borehole. This is a foreign key to DIC_DRILLING_METHOD
CLIENT_CODE	Client that instigated the request. This is a foreign key to DIC_COMPANY
PURPOSE_CODE	Purpose of the borehole. This is a foreign key to DIC_PURPOSE
COMMENTS	General comments or additional information not covered by other fields
X	The original X coordinate.
Y	The original Y coordinate.
EPSG_CODE	The CRS code for the original coordinate system. This is constrained against DIC_CRS_HORIZ_CS.
XA	The estimated accuracy of the X value. Example: If X is quoted as 302630 (metres) and known to within 5 metres, XA is 5 Example: If X is quoted as 6.1834528 (degrees) and known to the nearest second, XA is 0.0002778 (i.e. 1/3600)
YA	The estimated accuracy of the Y value. Example: If Y is quoted as 302630 (metres) and known to within 5 metres, YA is 5 Example: If Y is quoted as 6.1834528 (degrees) and known to the nearest second, YA is 0.0002778 (i.e. 1/3600)

BHD_LOG: A table within the Borehole database that holds geological information of the borehole log.

Column Name	Description
BHD_LOG_ID	Unique identifier of the borehole log populated by a sequence. This is a primary key of this table.
DRILLED_DEPTH_TOP	Depth at the top of the log.
DRILLED_DEPTH_BASE	Depth at the base of the log.
ORIGINAL_UNIT_CODE	Identifies the unit of measurement in which depths were recorded originally.
DEPTH_RELIABILITY_CODE	Reliability of the drilled depth. This is a foreign key to DIC_RELIABILITY.
LITHOLOGY_CODE	Lithology of the borehole record. This is a foreign key to DIC_ROCK_ALL.
LITHOSTRAT_CODE	Lithostratography of the borehole record. This is a foreign key to LXN_UNIT
UNIT_DESCRIPTION	Description of the unit
BASE_BED_CODE	Base of the bed of the borehole. This is a foreign key to DIC_BASE_BED.
BHD_INDEX_ID	Identifier of the borehole. This is a foreign key to BHD_INDEX

COMMENTS	General comments or additional information not covered by other fields.
BHD_LOG_SOURCE_ID	Identifier of the log source. This is a foreign key to BHD_LOG_SOURCE.
LAYER_ID	Order of the layer in the complete borehole log. This is also identifier of the individual log in a particular borehole.

BHD_LOG_PROPERTY: A table within the Borehole database that holds the borehole log property data.

Column Name	Description
BHD_LOG _ID	Identifier of the borehole log for which the property is recorded. This is a foreign key to BHD_LOG.
PROPERTY_TYPE_CODE	Identifier of the property type. This is a foreign key to DIC_PROPERTY_TYPE.
PROPERTY_VALUE	The value of the property type.
QUALIFIER	Qualifier for a property value such as : '?' - uncertain data '>' - greater than '<' - less than '>=' - greater than or equal to '<=' - less than or equal to
COMMENTS	General comments or additional information not covered by other fields.

BHD_LOG_SOURCE: Grouping table for logs (incl. interpreter, project, medium)

Column Name	Description
BHD_LOG_SOURCE_ID	Unique identifier of the borehole log populated by a sequence. This is a primary key of this table.
INTERPRETER_ID	Identifier of the person who coined the interpretation. This is a foreign key to STAFF table.
PROJECT_ID	Identifier of the project under which interpretation was carried out. This is a foreign key to PROJECT table.
SOURCE_MEDIUM_CODE	Identifier of the source from where the data was obtained originally. This is a foreign key to DIC_SOURCE_MEDIUM dictionary.

LXN_UNIT**:** The main table of lexicon database containing mostly index level information about the Lexicon Unit

** If there is a local alternative to the British Lexicon please use that instead.*

Column Name	Description
LXN_CODE	A Code to uniquely identify the Lexicon Unit. Each Lexicon Unit has one and only one "Lexicon Code" value. This should be up to 5 alphanumeric characters. Whilst there are many arguments for non-mnemonic codes the current codes are so embedded in other systems the cost of changing the systems would be high. However, business rules need defining to say that a "Lexicon Code" does not become obsolete just because a Lexicon Unit changes its name.
UNIT_NAME	The current name of the Lexicon Unit. This should be mixed case and be just as the name is normally written. Automated routines should be used to change existing entries (which are all upper case) and over 90% will be changed correctly; however, each will need checking individually. E.g. Earl David's Parrot Coal Allt a' Chuirn Quartzite Member This is a change from the current (25 Jan 2008) system where all names are entered in upper case. Business rules should be defined as to whether title case is to be used (for non-proper names) and whether the rank of the Lexicon Unit is to be included. It should not include an indication of whether the entry is obsolete or not.
UNIT_RANK_CODE	The codified rank of the Lexicon Unit. Each Lexicon Unit has one and only one "Unit Rank" value. It must use a controlled vocabulary including unclassified and include the "ranks" of non lithostratigraphical units.
STATUS	Whether the Lexicon Entry is obsolete or not. This should be O (for Obsolete) or C (for Current).
THEME_CODE	This is the DiGMapGB theme. All mapped rock units appearing in DiGMapGB belong to one of four themes: Artificial (including all anthropogenic deposits and excavations), Superficial (comprising deposits of Quaternary age except for the crag formations of East Anglia), Mass Movement (comprising landslide deposits), Bedrock (comprising the crag formations of East Anglia and all pre-Quaternary formations).

	Units appearing in the Lexicon but not in DiGMapGB are assigned to the theme in which they would appear, were they in DiGMapGB. Each Lexicon Unit has one and only one "BGS Theme" value. It should use a codified controlled vocabulary.
DEFNSTATUS_CODE	This is the state of the Lexicon Definition for a Lexicon Unit. It should use a codified controlled vocabulary of: Index Level Pending upgrade Full, pending ratification Full, without ratification Full It should not describe anything to do with the nature of the Lexicon Unit. Only the above controlled vocabulary should be available for new entries or changed definitions. However, old entries with ill defined entries, mixing "Status of Definition" with the nature of the Lexicon Unit ("Formal or Informal Unit"), "Definition Scope", where the entry is obsolete or not ("Obsolete Flag") and "Distribution Applicability" do exist and these need preserving to some extent.
FORMAL_OR_INFORMAL	This is the nature of the Lexicon Unit. Each Lexicon Unit has one and only one "Formal or Informal Unit" value. This should be Formal or Informal. This refers to the status and usage of the named rock unit. A formal unit is one that has been named according to the recommendations of the North American Stratigraphic Code (or as locally modified for BGS use), whether or not the definition has been completed or approved. All other units are informal. Units which could be given a valid definition, given sufficient information, should be designated "Formal". Units which cannot be given a valid definition should be designated "Informal". Some "informal" definitions are a descriptive expedient (e.g. the Upper and Lower Lincolnshire Limestone members are treated as informal divisions of the Lincolnshire Limestone Formation rather than formal ones because they are invalid in NASC terms). The state of completeness of a definition is dealt with by "Definition Status" (q.v.). This is a change from the current (15 Jan 2007) use of the Formal/Informal Flag which refers to the Status of the Definition.
ON_OR_OFFSHORE	This refers to the context in which the unit name is generally used. Offshore units are those whose names are generally used below the low water mark. They were typically recognised in boreholes or seismic surveys. They may extend onshore, perhaps with a different name. Onshore units are those whose names are generally used above low water mark. They may extend offshore, perhaps with a different name. Units designated as "Both" are recognised both onshore and offshore, with the same name. The definition of these units should preferably include both onshore and offshore reference sections. If there is doubt over designating a unit as "Both", one of the other designations should be chosen. If it is difficult to choose either "Onshore" or "Offshore", then the unit should probably be designated as "Both".
UNITCLASS_CODE	This is the "type" of unit, in a broad sense. It is analogous to the subdivision of rock types by origin into "igneous", "sedimentary" and "metamorphic", but uses broadly "stratigraphical" concepts. Each Lexicon Unit has one and only one "Lexicon Class" value. Most rock units will belong to one of the following classes: Man Made Ground Sedimentary - Lithostratigraphical Volcanic - Lithostratigraphical Intrusive Igneous Litho-morpho-genetic Unit Lithodemic - Metamorphic
PREFERRED_MA	The preferred map code for the Lexicon Unit.

PCODE	Each Lexicon Unit may have one and only one "Preferred Map Code" value. (If the Lexicon Unit does not have a "Preferred Map Code", a value of notApplicable should be used; values of notAvailable and notEntered should also be allowed.) This is not part of the definition of a unit but is useful information. Note that the "Lexicon Code" need not be the same as the "Preferred Map Code".
REVISION_NUMBER	Each Lexicon Definition has one and only one "Version Number". The first entry into the Lexicon (including the entry of index information) should be version 0. Subsequent versions of the Lexicon definition will be sequential integers. Minor changes should not be given new version numbers. Whether a new version number is given or not is at the discretion of the Lexicon Manager.
SCOPEBYAREA_CODE	Each Lexicon Unit definition may have one and only one "Definition Applicability" value. These terms do not refer to the geographical extent of the named units outcrop but to the extent to which the definition is applicable.

PROJECT: Table to store list of projects and their details.

Column Name	Description
PROJECT _ID	Identifier of the project. This is a primary key of the table.
PROJECT_NAME	Name of the project.
DESCRIPTION	Long Description of the project
PROJECT_TYPE_CODE	Type of the project. For example: Commercial, Science Budget, Enquiries and Sales etc. This can be constrained against a dictionary table.
START_DATE	Start Date of the project.
FINISH_DATE	Finish Date of the project.
PROJ_LEADER_ID	Project Leader assigned to the project.
STATUS	Current Status of the project. This has a check constraint in A: Active, C: Completed, I: Inactive.
PROG_ID	Identifier of the programme the project is part of.
BUDGET	Budget allocated to the project
DEPUTY_LEADER_ID	Identifier of the deputy Leader assigned to the project.

DIC_SOURCE_MEDIUM: Table to store the list of source types.

Column Name	Description
CODE	Code identifying the medium type used when creating an interpretation of the log. This is the primary key of this table
TRANSLATION	Short description of the medium type.
DESCRIPTION	Full description of the CODE value.
STATUS	Flag denoting the status of the CODE. This has a check constraint in C: Current, O: Obsolete.

STAFF: Table to store list of staff and their details.

Column Name	Description
STAFF_ID	Identifier of the staff. This is a primary key of this table.
FIRST_NAME	First Name of the staff.
LAST_NAME	Last Name of the staff.
NAME_PREFIX	Prefix for the Name.
BUSINESS_TITLE	Business Title of the staff.
JOB_CODE	Identifier of the staff job.
LOCATION	Location of the Staff in the organisation.
EMAIL_ID	Email Address of the staff.
STATUS_CODE	Employment Status of the staff.

Standard Dictionary Tables:

All of dictionary tables listed after DIC_BASE_BED have exactly the same structure as that shown below:

DIC_BASE_BED: A dictionary of base of bed codes.

Column Name	Description
CODE	Code identifying the base of the bed of the borehole. This is the primary key of this table.
DESCRIPTION	Full description of the CODE value.
TRANSLATION	A short translation of the code value
STATUS	Flag denoting the status of the CODE. This has a check constraint in C: Current, O: Obsolete.

DIC_COMPANY: A dictionary table to store list of client/company's details.

DIC_BOREHOLE_DIRECTION: A dictionary table to store start point types of the borehole.

DIC_CONFIDENTIALITY: A dictionary table of codes to describe confidentiality of a resource.

DIC_DATE_ACCURACY: A dictionary of valid Date Accuracies used to describe Dataset start/end dates of capture.

DIC_DRILLING_METHOD: A dictionary table of codes to describe drilling method of a borehole.

DIC_INCLINATION: A dictionary table of codes to describe inclination of a borehole.

DIC_ORD_DATUM_PREC: A dictionary table of codes to describe the accuracy of ordnance datum level.

DIC_PURPOSE: A dictionary table of codes to describe the drilling purposes.

DIC_XYSOURCE: A dictionary table of codes to describe the source of geographic coordinates.

DIC_RELIABILITY: A Dictionary table to store various depth reliability values for borehole geology.

DIC_ROCK_ALL: A Dictionary table to store all the different rocks.

DIC_PROPERTY_TYPE: Dictionary table to hold various property types for borehole geology.

Special Dictionary Tables:

The following dictionary tables have additional columns.

DIC_UNIT_OF_MEASURE: Dictionary table to store various measurement units.

Column Name	Description
CODE	Code identifying of the measurement unit. This is a primary key of this table.
DESCRIPTION	Full description of the CODE value.
TRANSLATION	A short translation of the CODE value
ABBREVIATION	Abbreviation code used to represent a measurement unit.
EPSG_CODE	EPSG code for a unit. This is a foreign key to DIC_CRS_HORIZ_CS.
DATA_SOURCE	Source of the data.
TARGET_UOM_CODE	Target Unit of Measure code.
CONV_FACTOR_NUMERATOR	Factor numerator to convert into Target code
CONV_FACTOR_DENOMINATOR	Factor denominator to convert into Target code
UNIT_OF_MEAS_DIMENSION	The dimension or feature for which this unit is used. For example length, pressure, temperature, angle etc.
REMARKS	General remarks or comments not covered by other fields
ORIG_INFORMATION_SOURCE	Original information source.
STATUS	Status of the unit. This has a check constraint in C: Current, O: Obsolete.

DIC_CRS_HORIZ_CS: Dictionary table of horizontal coordinate systems mostly sourced from EPSG database. Data from the normalised EPSG database is summarised and joined to populate this table. The table should be updated as new versions of the EPSG database are release.

Column Name	Description
CODE	Code identifying the horizontal coordinate system. This is a primary key of this table.

DESCRIPTION	Description of the coordinate system, e.g. OSGB 1936/ British National Grid
TRANSLATION	A short translation of the CODE value
DATA_SOURCE	The source of this coordinate system name and definition, e.g. EPSG v6.8, FGDC
EPSG_CODE	The EPSG code for this coordinate system, if applicable
COORD_SYS_TYPE	The type of coordinate system, inherited from EPSG database, e.g. projected, geographic 2D.
DATUM_CODE	The code for the horizontal datum from which the coordinate system is measured, can be used for look up in EPSG database or other non-EPSG database (not an internal foreign key).
DATUM_TRANS	The name of the datum of the geographic coordinate system, or of the geographic coordinate system that a projected system is based on e.g. OSGB 1936.
DATUM_ELLIPSOID	The name of the ellipsoid used by the geographic coordinate system datum, or by the geographic coordinate system datum that a projected system is based on , e.g. Airy 1830
SOURCE_GEOGCS_CODE	The code of the geographic coordinate system that a projected system is based on; points to another record in this table.
PRIME_MERIDIAN	The name of the prime meridian of the datum of the geographic coordinate system, or of the geographic coordinate system that a projected system is based on e.g. Greenwich
AREA_OF_USE_DETAILS	A description or list of the areas of the world in which the coordinate system is applicable, free text. Inherited from EPSG.
AREA_OF_USE_GROUP	The area of use group - WORLD, UK or UK-COMMON. This allows filtering and ordering of the dictionary to facilitate selection by users
COORD_TRF_EPSG_NAME	The EPSG name for the coordinate transformation for this projected system
UNIT_OF_MEASURE_CODE	The unit of measure that the coordinate system should be measured in, e.g. metres, degrees. Foreign key to DIC_UNIT_OF_MEASURE
STATUS	Current Status of the coordinate system. This has a check constraint in C: Current, O: Obsolete.